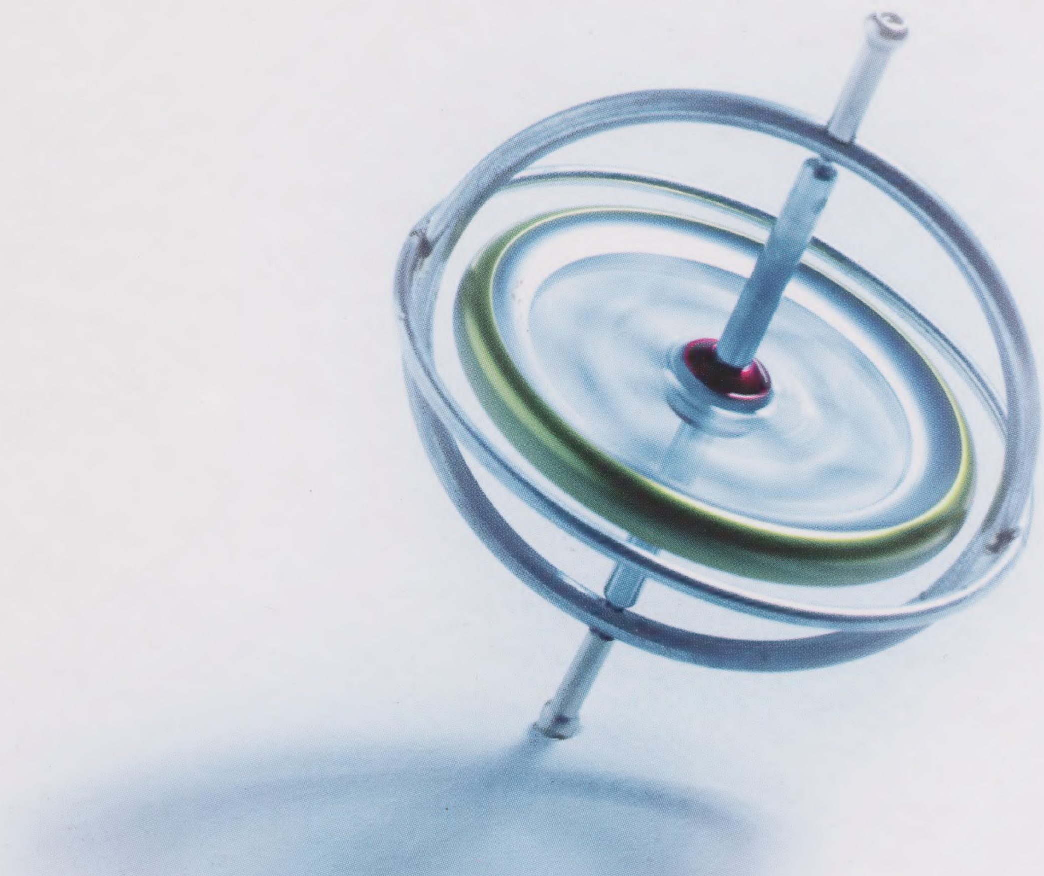
Cosmic Calculations

Astronomy and mathematics



Everything is mathematical

Cosmic Calculations

Cosmic Calculations

Astronomy and Mathematics

Rosa Maria Ros

Everything is mathematical

Published by RBA Coleccionables, S.A.
c/o Hothouse Developments Ltd
91 Brick Lane, London, E1 6QL

Localisation: Windmill Books Ltd.
Photography: Age Fotostock, iStockphoto, Corbis

ISSN: 2050-649X

Printed in Spain

Contents

Preface

One dictionary definition of astronomy is: "Science that pertains to the heavenly bodies and, mainly, to the laws governing their movements." The same dictionary's definition of "science" is: "Knowledge gathered by systematically structured observation and reasoning from which general principles and laws are deduced."

Be that as it may, it is quite clear that the language used in astronomy to systematise and structure the results of observation is mathematics, and it is a fact that mathematics has, on occasions, played a vital and fundamental role in the development of this science. But astronomy has particular characteristics. Astronomers will never be in a position to carry out and repeat experiments in the laboratory whenever they want, or be able to modify some conditions and then run the experiment again. It would be marvellous if we could have eclipses on demand, on whatever day we wanted and with the conditions of our choice. On a day with very few clouds, let's say. Or better still, with no clouds at all!

Astronomy was born along with humankind. As they had no TV or books, primitive peoples most likely spent the evenings gazing up at the sky or, at least, they will have gazed at it far more than any of us do today. That would be how they came to realise that some phenomena were repeated, and how these repeated events allowed them somehow to work out the best times to plant their seeds, and at what times of the year animals mated, meaning it would be easier to hunt them down. This no doubt had an influence on how they obtained their food and on their survival, and was, therefore, one of the first practical uses of scientific knowledge. Furthermore, certain inexplicable happenings they witnessed were 'interpreted' as the work of the gods, and so some phenomena came to be hallowed and led to ritual behaviour. The need to know and predict when these phenomena would occur was one of the purposes of the priests belonging to the diverse primitive creeds.

Astronomy has, in fact, always been very much in touch with ordinary people and, perhaps, even more so in the past than now. My grandfather was a farmer and I remember that he knew things that urban people are generally unaware of. For example, he told me that the Moon came out one hour later (really, it's 50 minutes, but for a farmer that's an hour) every night. My grandmother was well aware that the Sun was higher in summer than in winter – it came in through the window and reached further towards the end of the rooms depending on the time of the year. Although some of this knowledge has been lost, as far as the general public is concerned, it is a fact that some people, not necessarily professionals, have managed

to probe deeper into the science. I'm referring to amateur astronomers. It has always seemed to be very striking that astronomy is the science with the most amateur enthusiasts spread around the world. It is not at all unusual to come across people who, though their occupations may be totally unconnected with science, devote their spare time to scientific activities. There are, of course, people who spend their time gathering and classifying fossils or watching and studying birds, but the number of amateur astronomers' associations in the world easily outstrips that of amateur groups in any other scientific field. That is possibly because, even though some days it may be cloudy, we all have a sky over our heads, whereas fossils cannot be found just in any part of the world, nor are there birdwatching sites right outside our houses. Amateur astronomers are without doubt one of the defining characteristics of astronomy. That there should be so many of them is proof of a fact known by all science communicators – astronomy "comes across very well" and, as some of us say, "astronomy sells itself".

I strongly believe that astronomy comes across to the public more easily than other branches of science because it is very visual. It is very difficult to explain the latest findings in mathematics, for example, discoveries in number theory or differential geometry. On the other hand, it is very easy to show the latest images provided by a space telescope. Is there anyone who does not find some of their photos breathtaking? But, furthermore, if astronomy is well explained, it is exciting and hooks people just as a TV serial can. Who does not find it astonishing that stars are born, grow and die, and sometimes in dramatic fashion? Who can fail to be excited by the fact that it is in the entrails of the stars that the heavier elements that make up the atoms of our bodies are created? Who can fail to feel themselves part of the cosmos when they realise that we are simply children of the heavenly bodies, dust of stars? But more than that, the galaxies themselves grow and develop, and collide with each other. In short, astronomy is a world which is, in some senses, similar to the one we share with out neighbours. It is life itself but on an "astronomical" scale and provides us with truly beautiful images. How could it not be so popular?

People enjoy learning more about astronomy perhaps because in some way it tells us about the past, of the dawning of the Earth's spin, of the Solar System, of the cosmos, and therefore of our own home. It goes some way to explaining where we come from.

But astronomy also has another characteristic that defines it: prediction. It can predict some of the things that are going to happen – the sequence of the seasons, eclipses, the positions of the planets and stars in the sky and so forth. That

characteristic has at times been used by pseudosciences to try and predict things other than the behaviour of the skies. But that is perhaps an inevitable consequence of humankind's own characteristics. The feeling of insecurity regarding the future causes man to try to attain security by any means, and astrological predictions are aimed at that, though they have very little, if anything at all, to do with astronomy.

When it comes to predictions, there is a special relationship between astronomy and mathematics. Astronomical predictions are the fruits of mathematical calculation and, in fact, astronomy's needs have been met by the development of new mathematical fields. All of this well justifies the public's interest in discovering more about astronomy and, in particular, learning about the partnership between astronomy and mathematics.

I hope that readers enjoy this book, find answers to some questions that may be puzzling them and, why not, discover other questions that they want to delve into – that is the way of science. All researchers find themselves 'hooked' by two aspects: on the one hand, there is the almost indescribable feeling of having managed to understand or to solve something that they previously did not understand, and on the other, the 'bug' of another new question biting them. And so, in the same way that no scientists get bored when they are working in their field, I hope the reader likewise enjoys these pages. I really must say that for me it has been a pleasure to write them; I hope you get the same amount of pleasure from reading them.

The book consists of five chapters focusing on crucial themes in astronomy with mathematical content: positions and times, with the first two chapters dealing with positions and distances, and the last two – the fourth and the fifth – dealing with subjects related to time. The middle chapter, the third, is concerned with eclipses, those special situations of relative positions and coincidence in time.

Chapter 1

Crucial Angles and Required Distances: the ABC of Astronomy

It is obvious that a science dedicated to the observation and study of objects – albeit some certainly very special objects – must fundamentally be concerned with knowing exactly where they are. And in this primary aspect, mathematics plays a vital role. It carries out that role in the form of three values: two angles that will orientate us towards the celestial object and the distance the object is from us. Determining the two angles is relatively simple. However, to find out the distance to a heavenly body is one of the most problematic questions of all astronomy, and in this brief outline of the ABC of this exciting discipline we shall pay particular attention to that question.

Two angles for determining position

Coordinates are used to determine the position of an object on the surface of the Earth. As astronomical observations often depend on the position of the observer, as we shall see further on, their terrestrial coordinates are vital when processing astronomical data. We shall start the journey through the subtle art of astronomical positioning by going over how this method works.

Our planet revolves on an axis, the axis of rotation, which is normally used as a reference to fix points on the surface of the globe. For instance, the two intersection points of the axis with the surface of the planet are called the north and south poles. If we take the plane perpendicular to the rotation axis in the centre of the Earth we will see that plane intersecting with the axis forming an imaginary line that runs over the terrestrial surface: the Equator. This line divides the Earth into two hemispheres called the northern and the southern (at the tip of which are the two poles, north and south). If we imagine an infinite number of other planes parallel to the Equator and we make them intersect again with the

surface of the globe, we will get an equal number of circumferences, all smaller, which are known as parallels. Let's now imagine the planet as an orange divided into segments. Each segment is a great circle passing through the two poles and perpendicularly cutting the Equator. Let's give a name to those segments – how about meridians?

Unlike the Equator and the parallels, the meridians are all equal and, therefore, there is no obvious reason to distinguish between them. It was for that reason that in 1884 it was decided to take one of them as a reference – the 'zero, or prime, meridian' – and the one chosen was the meridian passing through the Royal Observatory at Greenwich in London. That meridian was not the first to receive such a distinction, but it is the one that has held it through to our times. In fact, until 1884, many European ships used El Hierro, the smallest of the Canary Islands, as their reference meridian, to be precise the line passing through Orchilla, on the western tip of the island. Ptolemy had chosen Hierro to mark the end of the known world and until 1492, nothing west of that point appeared on maps. It was, therefore, for many years considered a good place to have the prime meridian.

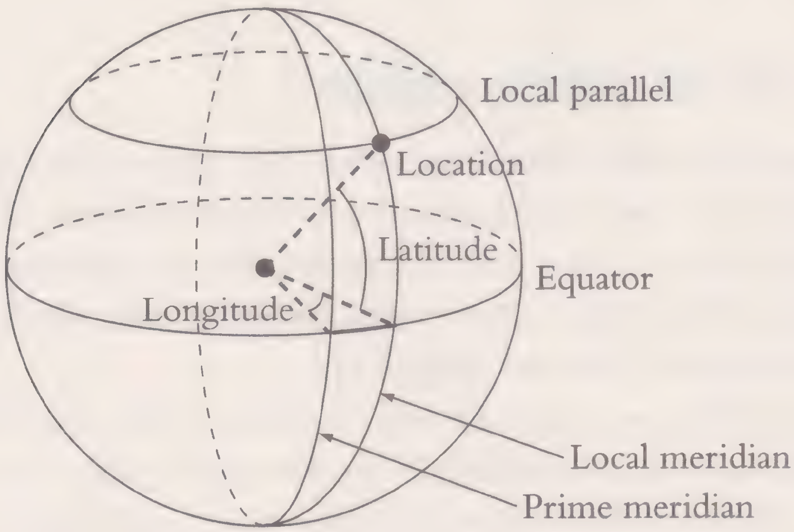


Diagram of the geographic coordinates, latitude and longitude.

We are now getting to the crux of the matter. Once the parallels and the meridians are established, locating a point on the terrestrial surface is as easy as establishing what parallel and meridian cross it. To do so we make use of two

ALICE'S ADVENTURES IN WONDERLAND

In Lewis Carroll's famous book, when Alice falls down a deep hole while chasing a rabbit, she has time to think about her situation (the text below is extracted from the beginning of the first chapter, entitled *Down the Rabbit Hole*):

> "Down, down, down. Would the fall *never* come to an end? 'I wonder how many miles I've fallen by this time,' she said aloud. 'I must be getting somewhere near the centre of the Earth. Let me see: that would be four thousand miles down, I think...' (for, you see, Alice had learnt several things of this sort in her lessons in the schoolroom, and though this was not a *very* good opportunity for showing off her knowledge, as there was no one to listen to her, still it was good practice to say it over.) 'Yes, that's about the right distance – but then I wonder what Latitude or Longitude I've got to?' (Alice had no idea what Latitude was, or Longitude either, but thought they were nice grand words to say.)"

Alice and the White Rabbit.

Without doubt, if the hole goes down, like any other normal hole, straight towards the centre of the Earth, it can be confirmed that her latitude and longitude will not vary. The angles of position are the same even though the distance to the centre of the Earth decreases. So Alice had no need to worry. Her latitude and longitude would not have changed throughout her fall towards the centre of the Earth.

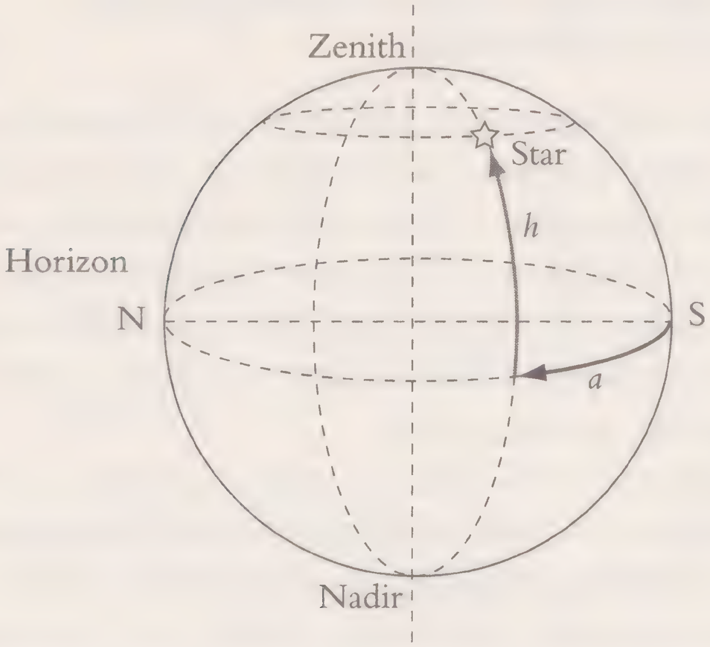
geographic coordinates, latitude and longitude. The first one is the angle measured along the local meridian or the meridian that runs over the place in question, from the Equator to that location. It is measured in degrees, from 0° to 90° for the northern hemisphere, and from 0° to –90° for the southern hemisphere. Longitude, on the other hand, is measured in degrees on the terrestrial equator, from zero meridian to the meridian of the location in question – from 0° to 180° towards the east, and from 0° to –180° to the west. The custom is to express terrestrial longitude in hours, minutes and seconds, instead of degrees. This other form of measuring longitude is particularly useful when it is used in the case of tools for the counting of time, such as for reading sundials, as we shall see later. To convert from angles to temporal units, it only has to be borne in mind that 24 hours correspond to 360° and that, therefore, one hour equates to 15°.

Localising bodies on the celestial sphere

In the past, humankind understood the sky as if it were a glass sphere placed over the Earth with stars attached to it. That concept has now passed into history, but we astronomers still speak of the celestial sphere partly because we are using lazy speech, but also because it does correspond with the intuitive view that one has of the firmament and the term can be useful in our work. The centre of this celestial sphere is the Earth, but not because it is considered as being the centre of the Universe, as it was in the days of Ptolemy, but because it is the place from where we are observing. Although nowadays we know that not all the stars are the same distance away from us, with the aim of clarifying the basic concepts we are dealing with, for the time being we shall assume (albeit reluctantly) that they are situated on the surface of the celestial sphere. The best way to pinpoint an object on a sphere is by taking some circumferences as a reference and by using a pair of angles from them, in a similar way to how we have done for locations situated on the terrestrial surface.

The first option: height and azimuth

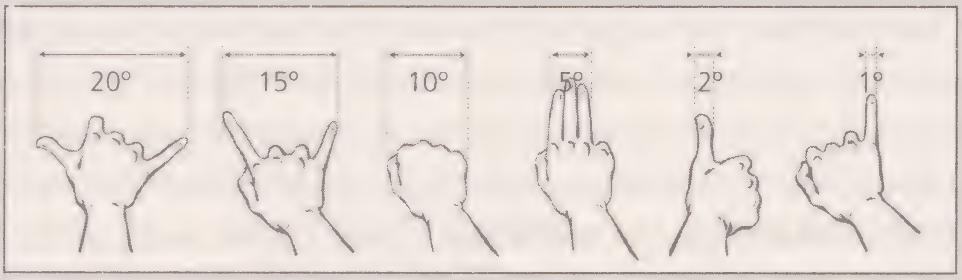
The most intuitive way to define coordinates is to give the angle from the observer corresponding to the height of the star over the horizon, and the angle from the line running north-south to the point on the horizon that appears to be directly beneath the star, ie, what is known as the azimuth (see the diagram opposite).



Height (h) and azimuth (a). In Europe the starting point of the azimuth is taken from the south, as shown, while in North America, it is taken from the north. The zenith corresponds to the intersection of the observer's vertical and the celestial sphere – in other words, the highest point in the sky – while the nadir is the opposite point.

MEASURING ANGLES WITH THE HANDS

A device called a theodolite is used for measuring the height and azimuth of a celestial body, but there is a very simple way to measure angles just by using your hands. Although it is not very exact, it is a very simple exercise that even children can do. If we fully stretch out an arm in front of us, the span of the hand corresponds to 20°, the fist to 10°, the thumb to 2° and the little finger to 1°. Whether the person doing this is large or just a child, as their bodies are proportional, their arm corresponds with the size of their hand, and the rule gives good results.



How to use a hand, held at arm's length, to gauge angles.

The second (and better) option: declination and right ascension or hour-angle

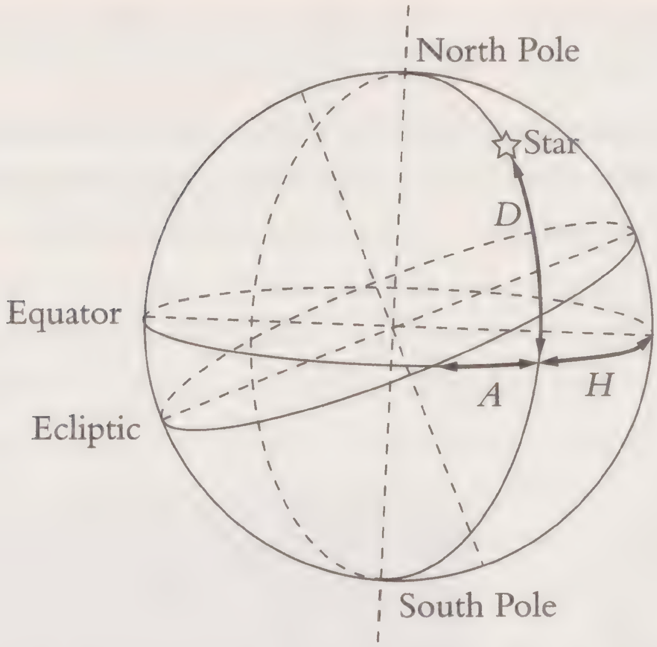
The determination of the position of a star by height and azimuth is a simple method, but it has one serious drawback: it is very local. The same star observed simultaneously from Paris and London will have different coordinates, as the horizons of the two places are different. For instance, the North Star is seen as higher over the horizon from London than it is from Paris. These coordinates, as simple as they are, do not allow astronomers in different places to exchange location information on the observations they are carrying out.

There is another way of giving the angles of position which also has some features in common with the terrestrial coordinates of longitude and latitude and which solves the problem of simultaneity. The method is based on the notion of the Earth's axis of rotation and it is also intuitive for us because it responds to the obvious perception that it is the celestial sphere that is rotating around us, while the truth is precisely the opposite: what is rotating is the Earth, from west to east. But, because we are 'riding' on the Earth, we perceive the inverse effect.

Let's therefore look at the plane that cuts the celestial sphere perpendicular to the axis of rotation passing through the centre of the Earth and the celestial sphere. This plane cuts the terrestrial surface in a great circle which is known as the terrestrial equator, and cuts the celestial sphere in a great circle called the celestial equator. The second analogy will be that of a celestial meridian that passes through the two poles on a plane perpendicular to the Equator.

As all celestial meridians are equal, the same as on the Earth, one is taken arbitrarily; in this case the meridian that passes through the point where the Sun is situated on the first day of spring, known as the Aries point. The position of any star or astronomical object is then defined by two angles: declination and right ascension, as can be seen in diagram on page 17.

The declination is the angle from the Equator to the star on the location's meridian (it can be from 0° to +90° or from 0° to -90°); the right ascension is the angle on the celestial equator from the Aries point to the star's meridian (from 0 to 24 hours). Sometimes the hour angle is used instead of right ascension. The hour angle, on the celestial sphere, is the angle of the celestial body's position with respect to the celestial meridian of the observation point (or local meridian).

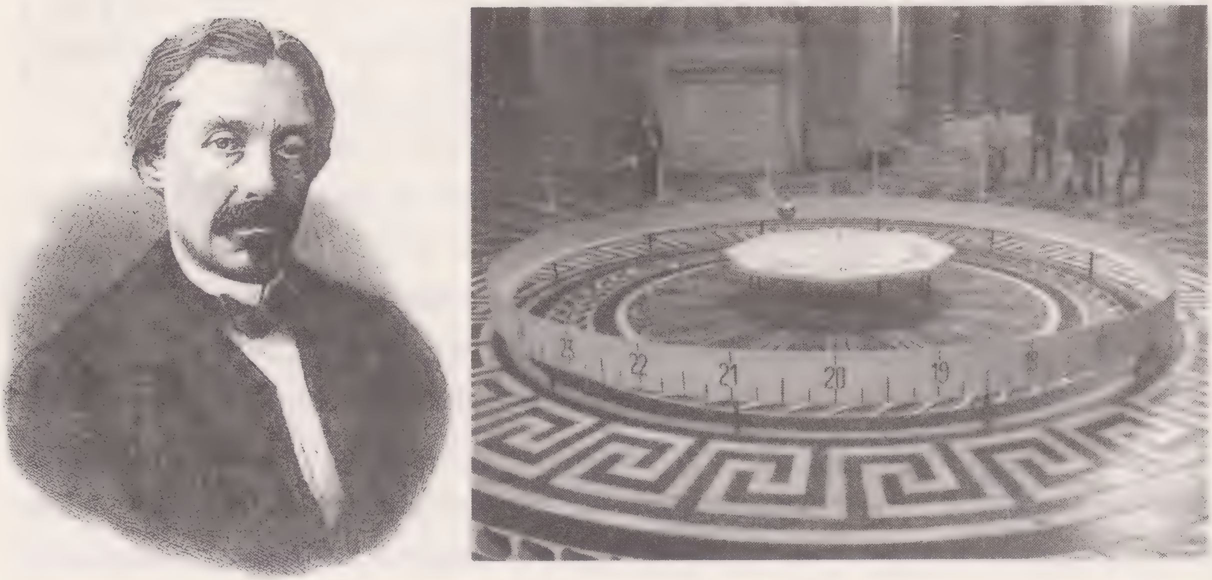


The position of a star using declination and right ascension (A,D) and hour angle (H,D).

The advantage of the equatorial coordinates system – declination and right ascension – over the previous one is very clear – it is universal. The coordinates are the same for any object and for all observers. By definition, moreover, they are based on the nature of the planet's rotational motion, which allows corrections to be made to the distortions in that motion with respect to our astronomical observation. And the fact is, as was mentioned above, the celestial sphere's apparent rotational motion is really an effect relative to the rotational motion of the Earth. It is the same sensation felt when we are sitting on a train at a station and we see the train on the track beside us move. We cannot know which of the two is moving by observing just the two of them. To find out, another reference is needed: in this case it would suffice to look out through the opposite window at the station platform to see if it was 'stationary'. However, if, instead of the trains, we consider the case of the Earth and the celestial sphere, that additional reference is not so simple to find.

In 1851, the French scientist Jean Bernard Léon Foucault (1819-1868) devised an experiment that demonstrated the movement of the planet in relation to the celestial sphere. Foucault hung a 28kg mass from the dome of the Paris Pantheon on a 67-metre length of cable. Foucault's pendulum swung for 6 hours with a period of 16.5 seconds, and it deviated 11% per hour. In other words, as time went by, the pendulum's oscillation plane moved with respect to the building. However, it is known that pendulums always move on the same plane of motion. (We suggest that the reader tries it. Simply hang a key ring from a thread and let it swing; it will

always swing on the same plane). The deviation recorded, therefore, could only be from one cause. It was the building, and by logical extension, the Earth that was rotating around the pendulum's oscillation plane. This was the first objective proof of something that was apparently so elementary – the Earth rotates. There are now pendulums of this type set up in many cities and they are a common attraction in science museums.



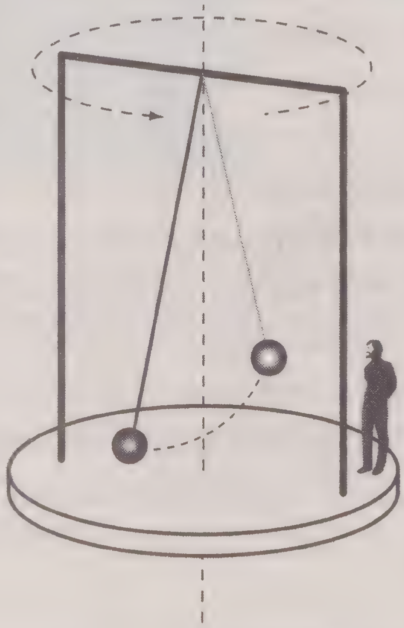
Jean Bernard Léon Foucault and a view of the pendulum at the Paris Pantheon.

It is, in fact, a good thing to bear in mind that the Earth, which intuitively seems to us to be so still, not only spins on its axis, completing a rotation every 24 hours (which is equivalent to a speed of 1,600km/h, in other words, about 500 metres per second at the Equator), but also revolves around the Sun, making a complete revolution in 365.2522 days (at an average speed of 30 kilometres per second, that is, 108,000km/h). But, furthermore, the Sun is moving with respect to the centre of our galaxy making one complete revolution every 200 million years and moving at a speed of 250 kilometres per second (900,000km/h). And, on top of all that, our whole galaxy, is moving away from all the others. It is as if we were on a fairground ride. Not just any ride though, but on one of the wildest. We rotate, move forward and are launched through space in a spiral at a dizzy speed. But to us it seems like we're standing still!

MAKE YOUR OWN FOUCAULT PENDULUM

We suggest the reader tries this simple experiment. Take a round tin or jar and stick a thick piece of cardboard or a thin sheet of wood on the lid on top of which you fit a little frame, as in the diagram. You can simply use two vertical sticks with a horizontal one placed on top, like a door frame. On the edge of the base, fix a little figure of a person who will be your 'observer'. On the horizontal bar of the frame, fix a thread with a weight. Start the pendulum off so that it swings parallel to one of the walls in the room you are in. If you gently turn the platform via the round lid, you will see the frame and the little human figure change position in respect to the reference wall, but the plane of the pendulum's movement will continue to be parallel to the wall.

If we now imagine ourselves in the place of the little figure of the observer, we shall see that his or her perception is that the pendulum moves with respect to the floor, as the observer does not perceive his or her own movement and that of the frame. Likewise, in a museum, when we watch how a pendulum moves with respect to the building, we get the feeling that the pendulum is moving its position (very gradually), but in reality it is us, together with the building and the whole planet Earth, that are on the move.



Although there are other types of coordinates in astronomy, the two systems that we have explained so far are, as well as being the most interesting, more than enough for anyone who wishes to understand some astronomical aspects but does not want to specialise in them. There is just one last question left to be dealt with. How can we make observations carried out with one system compatible with those of the other system? Readers interested in this will find the answers in the Appendix.

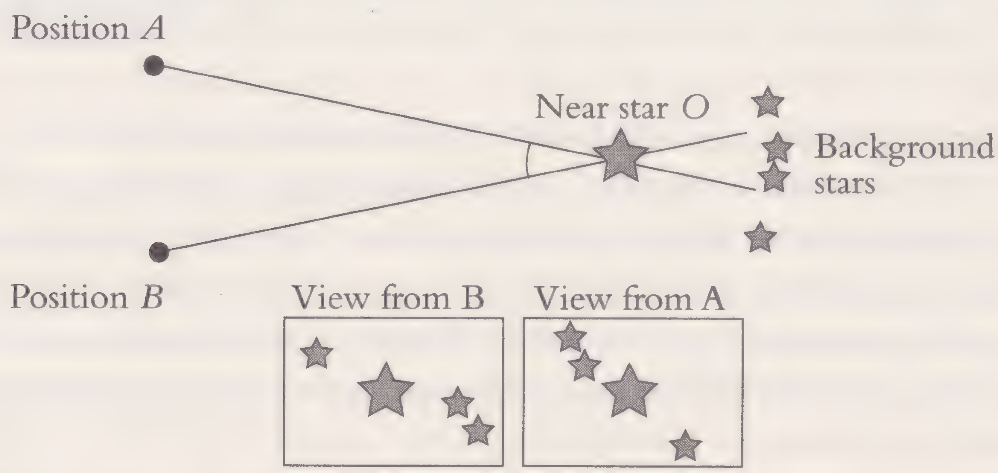
The problem of distance

Determining the positioning angles of any astronomical object is relatively simple. Basically, it is nothing more than playing 'battleships' – a system of two coordinates.

The real problem, as we mentioned at the beginning of the chapter, lies in determining the distance of the astronomical object that we want to observe. There are specific methods for determining distances that depend on the physical properties of the objects. As our subject here is the mathematics of astronomy, we shall not go into them, and we shall just give a method applicable to different objects which is often used: the determination of distance by parallax.

Parallax measures the deviation obtained in the position of an object in respect to a reference point when the observer changes their position or location. We have all experienced this phenomenon. If we take a photo with a camera other than an SLR camera – in other words, we take the photo with a camera that has a viewfinder situated in a different place than the lens, sometimes we are disappointed to find that the framing that we saw on taking the photo does not coincide with how the photo comes out. Someone on the edge of the photo may have been cut off, or a person's feet may have disappeared. This is because we do not see exactly the same thing through the viewfinder as the lens 'sees'.

We car drivers can also experience something similar when reversing and we look round by turning our heads: what we see when we look round from the right is not the same as when we look round from the left. Let's suppose there is a lamppost on the pavement. If we look round from the right, we see it at a certain point against, for example, the front of a house, while if we do the same from the left, we will see it displaced in respect to that same background.

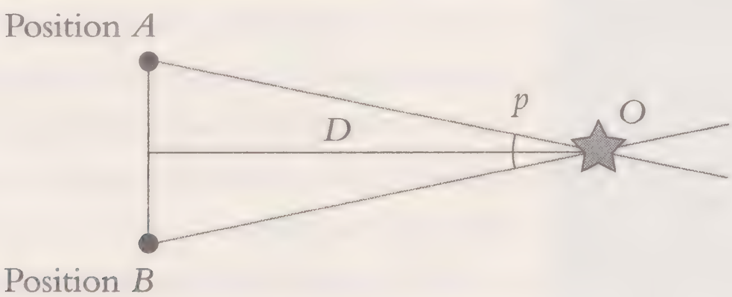


Let's go on then to look at the parallax effect and to see its applications in the field of astronomy. As the diagram above shows, the position of the nearby star O varies according to the position from where it is observed. When the image of the

star on the starry background is projected far enough it can be seen to change position – from A, star O seems to be on the left of the pair of stars, while from B it seems to be on the right of them. The angle covering the segment AB from O is the angle of parallax. This angle is always very small, particularly for objects outside the Solar System.

Two photographs of the Moon taken (one from the UK and one from Canada) on 28 October 2004 laid one on top of the other so that the background stars lay perfectly on top of each other. The apparent change of position of the Moon is due to parallax (source: Pete Lawrence and Peter Cleary).

For example, if we observe the Moon with respect to its starry background, we can calculate how far away it is provided that we know the distance between the two places on the Earth. Let's take the following figure:



From basic trigonometry, we know that:

$$\tan\frac{p}{2} = \frac{\left(\dfrac{AB}{2}\right)}{D},$$

and therefore, to calculate distances, we have:

$$D = \frac{AB/2}{\tan p/2} = \frac{AB}{p},$$

where the tangent to the angle has been approximated, as the angles are very small.

Different types of parallax can be defined. Let's go back to the previous diagram. If we take positions A and B as those occupied by the Earth at the two moments that it is furthest away from the Sun, we will have what is known as the annual parallax. The base of the triangle is, therefore, the distance that separates one position from the other, ie double the Earth-Sun distance, some 300 million kilometres. The 150 million kilometres separating the Earth from the Sun is known as the astronomical unit (AU) and it is often used to measure large distances within the Solar System. By using this value, once the parallax angle p is deduced, we can deduce that the distance to the star is, in kilometres, $d = 300{,}000{,}000/p$, where angle p must be expressed in radians.

Observing parallax with just one finger

This is a very simple exercise that can be carried out by looking at the position of your finger relative to a background, such as a wall. To do it, raise your right arm with the index finger upright (as shown in the photograph); close your left eye and note the position that your finger has against the background. Without moving either your arm or your finger, open your left eye and close the right one, and again observe the position of your finger in respect to the background wall. You will see your finger change position depending on the eye you are using. This is exactly the same phenomenon that is used in astronomy, and the only factor that differs is the scale.

Observing the parallax phenomenon with one finger.

It is precisely these observations made by each of the eyes that enable us to see things in relief and enables our brains to make an estimate of the distance separating us from the object. We can put this ability to the test thanks to a simple experiment. In some souvenir shops, it is possible to buy a simple device made of cardboard which appears to have the same

photograph twice over. In actual fact that is not the case; the photos were taken from two different points a few centimetres from each other. When you look through the two holes, one for each eye, your brain merges the two images into a third which you see in relief. In fact, these devices, called stereoscopes and which were considered to be amusements in the 19th century, simply make use of the parallax effect.

A stereoscope shows two similar photos which are viewed through the eyepiece. Our brain merges the two images into one in which we can see the relief that is lacking in each photo when viewed separately.

3D cinematography is based on the same principle. The movie is recorded simultaneously by two lenses separated by a certain distance, and in the cinema the two images are projected simultaneously. To see the film in 3D, viewers have to wear glasses which can be of different types, but in all cases allow the observer to see one of the images projected with one eye, and the other image with the other eye. When the brain puts the two images together, we get the sensation that we are seeing things in three dimensions. To get the optical illusion of 3D there are several possibilities. One type of glasses has a red filter for one eye and a blue filter for the other; the movie is filmed in such a way that, depending on the filter used, we will only see one part or the other of the simultaneous projection. With another type of glasses, the lenses are polarised, with a different type of polarisation in each, also with the aim of the viewer being able to see one image or the other with each eye.

The determination of parallax

Parallax has given rise to the use of a unit of measurement called the parsec (from the words parallax and second and often abbreviated to pc). A parsec is the distance

WHAT DO YOU NEED TO KNOW BEFORE BUYING A TELESCOPE?

All telescopes comprise two parts: the mount and the lenses. We shall not deal with the lenses here, but we shall briefly explain the different mounts for supporting telescopes. That way, readers who are inexperienced in this field will have some understanding if they ever go to a shop to buy one for themselves. For each system of coordinates there is the right kind of mount, which offers a different kind of service from the others.

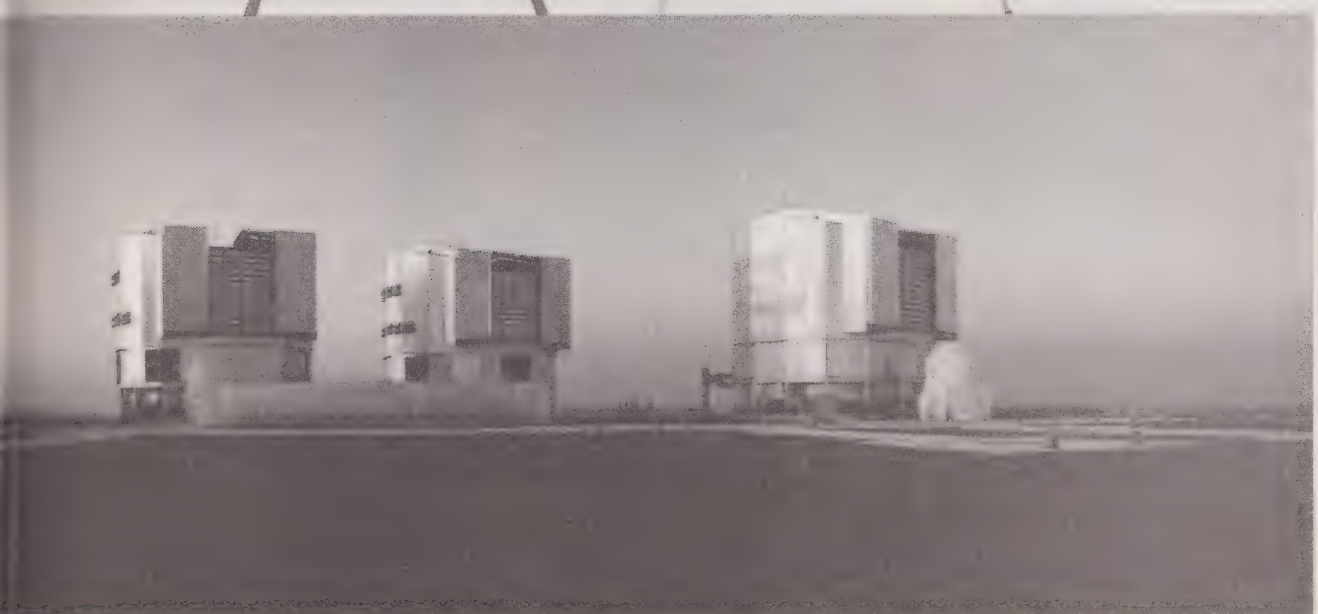
Telescopes with altazimuth mounts have more stability than those with equatorial mounts, but the mount itself makes observation more difficult as it is not an easy task to make small corrections for the celestial sphere's rotational motion. When tracking any object's trajectory with one of these telescopes, the height and azimuth have to be moved so as not to lose the object. This can turn out to be complicated. Their advantage is that they are much cheaper and are very simple to set up – they are no more difficult to set up than the tripod for a camera. They can be put up in any fashion and wherever you want.

Telescopes with equatorial mounts are more unstable, so the counterweight has to be set very carefully, and, on top of that, they are more expensive; you can see why just by looking at the different structure of the mount, as the equatorials are much more elaborate. They have the drawback that to use them they have to be set up in such a way that the telescope's main axis is oriented in accordance with the Earth's axis of rotation. Their great advantage is that corrections

A view of the array of telescopes that make up the VLT (Very Large Telescope) at Cerro Paranal, in Chile. Very large telescopes employ azimuth mounts because of their greater stability. They have no problems tracking because they are equipped with computers which dictate

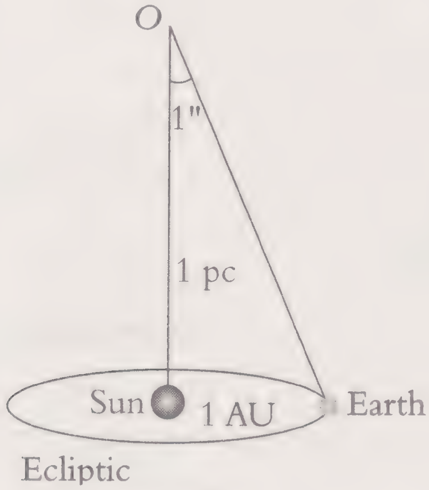
for the rotation motion can be made by moving it slightly in right ascension, so it is easy to install a small motor to operate this tracking process. This type of mount is definitely the most attractive to the prospective amateur astronomer.

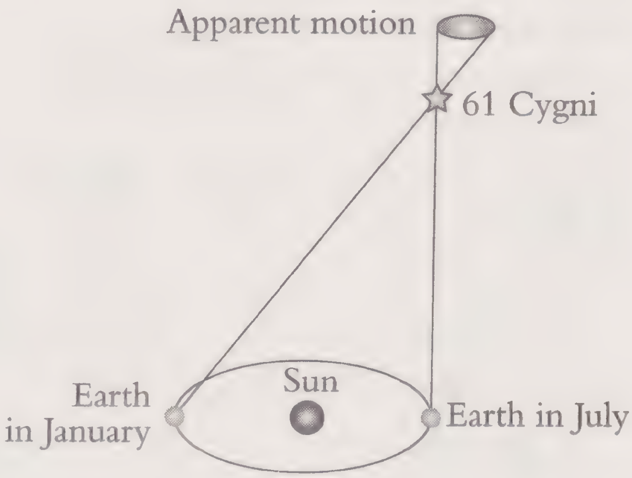
Two telescopes with different mounts for amateur astronomers: on the left, a telescope with an equatorial mount to which a motor drive for tracking can be fitted; on the right, a telescope with an altazimuth mount.



the movements they make and take care of the necessary corrections. Nowadays, in fact, astronomers observe the sky on the screen of a computer, and so the romantic image of a professional astronomer peering into an eyepiece no longer corresponds with reality.

that produces a parallax with an angle of one second of arc when viewed from two locations separated by 1 AU (150 million kilometres). It is equal to 30.9 trillion kilometres, or 3.26 light years.





Above left: the parsec (pc) corresponds to the parallax of one second of arc.
Above right: the annual parallax of the star 61 Cygni.

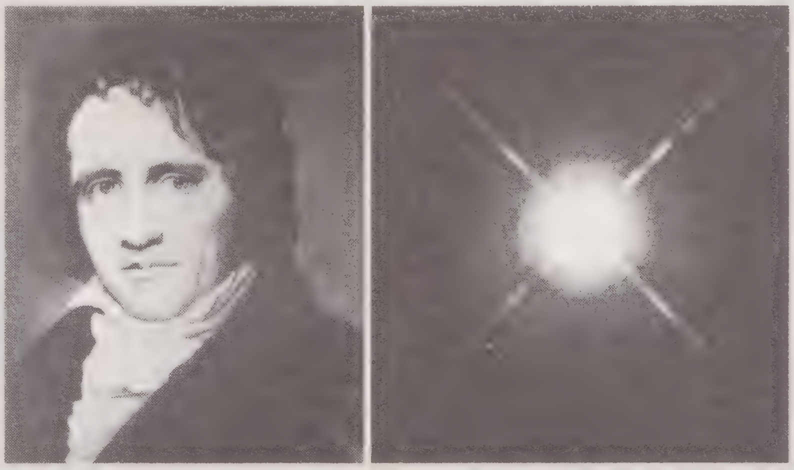
This unit is very widely used in astronomy, as are its multiples: the kiloparsec (a thousand parsecs), for measuring distances on a galaxy scale, and the megaparsec (a million parsecs), for measuring intergalactic distances (though these distances are too great for a true parallax to be seen). The distances separating us from the stars, even the nearest ones, are so huge that meticulous telescopic observations have to be carried out to be able to find their parallax. The ancient Greek astronomers knew about parallax, but as they lacked instruments precise enough to observe annual parallax, they came to the conclusion that the Earth stayed still relative to the Sun.

The first person to determine the parallax of a star – 61 Cygni, in the Cygnus constellation – was the German astronomer and mathematician Friedrich Wilhelm Bessel in 1838. To get an idea of how small the parallax can be, even of the nearest star, let's take the case of the stellar system that is closest to Earth, the one formed by the three stars of Alpha Centauri. Proxima Centauri, the nearest, is 40 trillion kilometres away, i.e., 4.3 light years. This star has, therefore, a parallax of less than one arc-second, 0.765, which is less than 1/3600th of a degree, or, put another way, 1/3600th of the angle covered by our little finger with our arm outstretched. In other words, it is incredibly small.

The greater the distance, the smaller the parallax, and any errors made become more significant. From 100 light years upwards, the trigonometric annual parallax is not reliable for determining stellar distances and other methods have to be used.

FRIEDRICH WILHELM BESSEL (1784–1846

Bessel, a German mathematician and astronomer born in Minden and who became director of the Königsberg Observatory, systematised the Bessel functions (discovered by Bernouilli) and worked on the calculation of orbits, measurements of stellar positions, aberration and atmospheric refraction. His interest in astronomy and navigation led him to work on the resolution of spherical polygons, and to create the famous Bessel formulae for resolving, among other things, the pole-zenith-star triangle. He achieved a high level of precision with his instruments, which enabled him in 1838 to determine the annual parallax of the star 61 Cygni, after 18 months' observation. In 1844, by analysing the micrometric positions of Sirius and Procyon, he proved that their movement in the sky could only be explained by the presence of an invisible body which was affecting their orbit. Using this observation, he deduced the orbit of Sirius B, which would not be discovered until 1862, and that of a partner for Procyon, not discovered until 1895. Bessel is also known for having published a catalogue with the exact positions of 75,000 northern hemisphere stars.



Above left: Friedrich Bessel. Above right: a photograph of Sirius A (the large star) and Sirius B (the star barely visible at the 7 o'clock position) taken by the Hubble Space Telescope (source: H. Bond, STScI, and M. Barstow, University of Leicester; NASA, ESA).

The planet hunters

As well as the parsec mentioned above, there is another useful measure: the light year (ly), in other words, the distance light travels in a vacuum in one year. As light travels at 300,000 kilometres per second, a light year corresponds to 9.46 million million kilometres, that is, getting on for 10 trillion kilometres. When the field of reference

is the Solar System, the unit of distance used is, as we saw above, the astronomical unit (AU), ie 150 million kilometres. To be able to work and study in the simplest way in astronomy, it is important to use the units of distance that are most appropriate for each case. We are certainly not able to imagine such magnitudes as billions of kilometres, it is much clearer to say that Jupiter's distance from the Sun is five times the Earth-Sun distance, in other words, 5 AU, or that Saturn is 10 AU away.

The choice of suitable units of measurement is important as it helps simplify research and enables the corresponding results to be interpreted better and more easily. The perfect example of this is the Titius-Bode Law. This law postulated a fixed relationship between the distances of the Solar System's different planets, which encouraged the 'hunting down' of new bodies which occupied the gaps predicted by the law.

The Titius-Bode Law established a relationship between distances in AU and the position that each planet occupies in the Solar System. The law was proposed in 1776 by Johann Daniel Titius, but it was for many years attributed to the director of the Berlin Astronomical Observatory, Johann Elert Bode, who was the one who popularised it best at that time.

Above left: Johann Daniel Titius.
Above right: Johann Elert Bode.

The law can be formulated as a succession whose general term verifies that:

$$a_n = 0.4 + 0.3 \cdot 2^{n-2} \text{ with } n = 2, 3, 4\ldots \text{ and for the case } n = 1\ a_1 = 0.4.$$

The law therefore predicted a succession of planets separated by the following distances:

n	Planet	Titius-Bode Distance	Real distance (AU)
1	Mercury	0.4	0.39
2	Venus	0.7	0.72
3	Earth	1.0	1.00
4	Mars	1.6	1.52
5	?	2.8	2.77
6	Jupiter	5.2	5.20
7	Saturn	10.0	9.54

Succession of planets devised at the end of the 18th century according to the Titius-Bode Law for the Solar System.

As you can see, at first sight the law works quite well. In its classical formulation, the law establishes a progression ratio of 2, but if the adjustment is made with precision it is seen to respond better to a ratio of 1.71.

The Titius-Bode law aroused tremendous excitement among mathematicians and astronomers. At that time, no planets other than those shown were known, and it is easy to imagine the interest among the scientific community in finding the planet situated between Mars and Jupiter, whose existence – and even distance! – was predicted by the law. Others searched beyond Saturn. And it was in 1781, not long after the law was formulated, that the British-German astronomer William Herschel discovered Uranus at 19.19 AU, just a few million kilometres further in than the predicted 19.6 AU. It hardly needs saying that this discovery provided huge support for the work of Titius and Bode.

At the 1796 astronomical congress held in Gotha, Germany, the eminent French astronomer Joseph Lalande urged his colleagues to go on searching for the nearer planet. In 1800, the German Franz Xaver von Zach organised a search party covering zones of the zodiac using a team of 24 astronomers, who came to be known as the 'celestial police'. Von Zach and his colleagues discovered a large number of asteroids, as they would later be called, but the jackpot prize would go to a person who had no connection with the group: the Italian Giuseppe Piazzi, who spotted the sought-after planet on 1 January 1801, and gave it the name Ceres. It was exactly 2.8 AU away.

After observing it for 24 days, Piazzi wrote a letter to Bode announcing the discovery. The German did not actually receive the letter until the end of March; by that time the object had moved so near to the Sun that it was impossible to see. Piazzi attempted to calculate its position, but there were only methods for calculating circular and parabolic orbits, so the Italian, who was certain the object he had seen was not a comet, calculated a circular orbit for it. When he started his search following the period of invisibility, Ceres was not in the place he had anticipated.

At that time, a young German mathematician named Carl Friedrich Gauss was developing a method specifically for the calculation of elliptic orbits from three measurements. In October 1801, Gauss received a letter from Von Zach with details of Piazzi's observations and explained to him the difficulties they were having recovering Ceres. Gauss applied his new method to the data he had been given and on 7 December 1801 Von Zach located Ceres exactly in the position predicted by Gauss, a long way from the region where they had been looking for it.

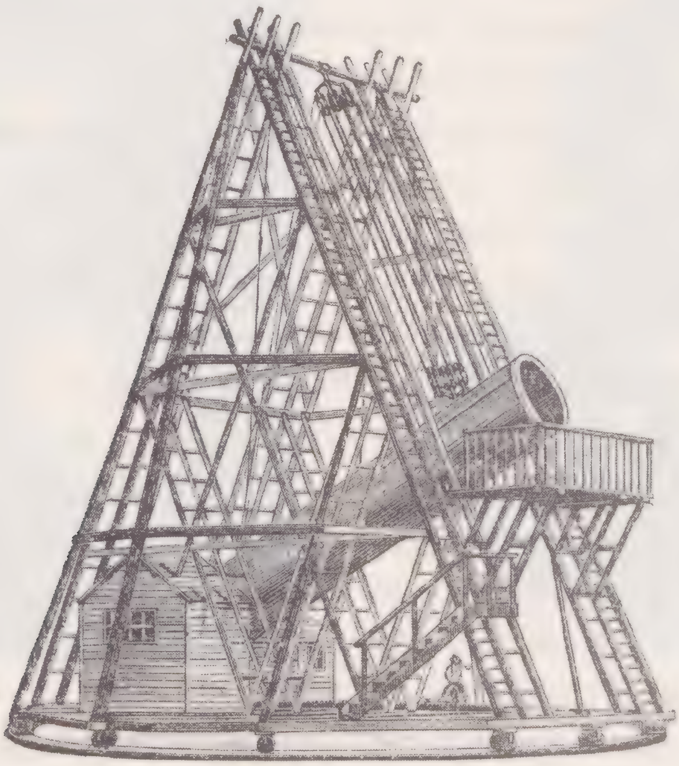
But as a planet, Ceres seemed suspect, since its measurements showed it to be smaller than the Moon. Furthermore, a year later, Heinrich Olbers, a compatriot of Gauss, discovered another object which he named Pallas, in a similar orbit, and in 1807 he detected another two: Vesta and Juno. All of them appeared to be planets, but on account of their small size, William Herschel theorised that they could not be, and he went on to call them asteroids. Due to the technical limitations of the telescopes available at that time, it was impossible to find more, and Ceres kept its place as the missing planet.

With the birth of astronomical photography, the situation completely changed and, by 1900, there were 436 known asteroids. Nowadays, we know that the asteroid belt, situated between the orbits of Mars and Jupiter, holds about 400,000 objects with a total mass of four per cent that of the Moon. They are not remnants from any planet destroyed by some cataclysm, as thought, but fragments of a planet whose creation process was halted. Ceres finally lost its status as a planet but has recently been elevated to more than just the largest of the asteroids, and is now one of the growing number of dwarf planets.

In 1846, Neptune was discovered at 30 AU, whereas the law had predicted 38.8 AU. Now, after more than a century had gone by, the Titius-Bode Law came to be considered simply a mathematical curiosity; a sad ending for what was one of the principle driving forces behind progress in astronomy at the end of the 18th century and the beginning of the 19th century.

Neptune: discovered with a pencil and paper

William Herschel had built the best telescopes of the era and devoted himself to watching the skies with such enthusiasm that it was only a matter of time before he came across Uranus, discovered in 1781. The discovery of Neptune, however, was not the fruit of an exhaustive search, but the result of mathematical research.

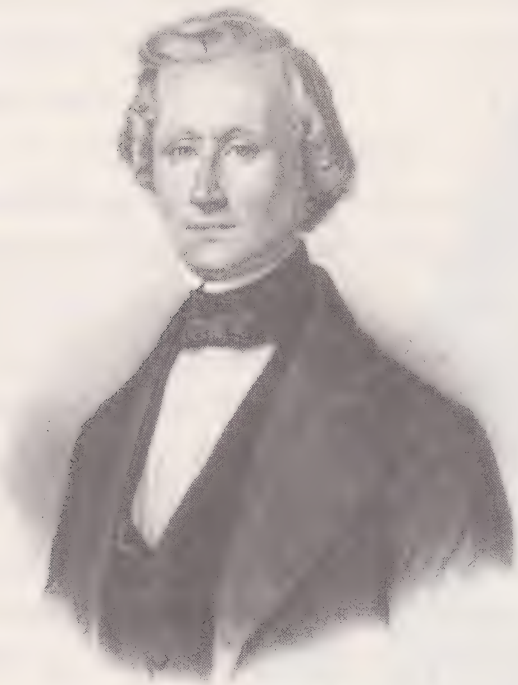


William Herschel's largest telescope had a 1.2m aperture. The British–German astronomer built it over two years, concluding the work in 1789.

Following Herschel's discovery, an in-depth study of Uranus' orbit followed, and tables were calculated indicating where the planet should be on any particular date. But as time went by it was seen that Uranus was not following the exact orbit calculated; there was some unknown cause for these alterations. But what? There was so much interest among the scientific community in solving the mystery that in 1842 the science academy at Göttingen offered a prize for anyone who could solve the problem of Uranus' movements. Soon two scientists predicted the existence of a new planet, albeit by different means.

The story of the discovery of Neptune would make a good television drama. It is a tale of men of science, with their personalities and characters, and of scientific research. Two mathematicians, Urbain Le Verrier, a renowned and highly regarded

Frenchman, and John Couch Adams, at that time a young and unknown Englishman, were analysing the slight deviations in the positions of Uranus and, working completely independently, both arrived at the same hypothesis: the deviations must be due to the gravitational pull of an unknown planet beyond Uranus. They both, therefore, predicted that there must be a planet in the same area.



Above left: Urbain Le Verrier. Above right: John Couch Adams. Today, the discovery of Neptune, the most distant planet in the Solar System, is attributed to both of them.

After finishing his studies at Cambridge University, Adams began working on the Uranus problem. In October 1843, he found the mathematical solution and asked George Biddell Airy, the Astronomer Royal, to provide him with more detailed data on the planet's motions so that he could make more precise calculations. In September 1845, Adams sent Airy the results of his calculations (very similar to those that Le Verrier would obtain a year later). Adams, as an unknown, did not manage to get Airy's full attention, no doubt due to a clash of personalities. Airy suggested that he should ask James Challis, the director of the Cambridge Observatory, to search for the new planet. Finally, Challis started the search and, in actual fact, he did observe Neptune, but because he was looking for changes in the position of any of the objects that he had recorded, he failed to notice that one of the brightest in the search field showed a small disk (not a mere point of light) – and was therefore the planet in question.

In September 1846, Le Verrier concluded his calculations and wrote to Johann Friedrich Galle at the Berlin Observatory, which had the best telescopes of the time, to ask him to carry out the relevant observations at the point in the sky where he predicted that the new planet must be. As soon as he received the request from Le Verrier, a well-known astronomer, Galle, got to work. Five days later, on 23 September, the planet was located, very close to the spot predicted.

It must have been even more frustrating for the unfortunate Adams when, in 1847, his own country's Royal Astronomical Society presented Le Verrier with an award for the calculations which had led to the discovery of Neptune. Fortunately, the following year, justice was done and the same society gave the same award to Adams. In 1861, he became the director of the Cambridge Observatory. Nowadays, recognition for the discovery is attributed equally to both men. Subsequently it was found that Galileo had already seen Neptune, but due to his low-powered telescope he took it for a star. His drawings of 28 December 1612, and 27 March 1613, show the planet as a star close to the position of Jupiter.

Uranus was discovered by observation, but finding Neptune was the fruit of the power and reach of mathematics. It was calculations that guided the telescope which, eventually, would confirm predictions made with pencil and paper and, although mathematicians tried to repeat the achievement, they did not have the same success. The discovery of Neptune has gone down in history as a triumph of celestial mechanics.

Chapter 2

But Where Are We?

Since the first humans appeared, it seems that they have always thought of themselves as being somehow the centre of the Universe. With the arrival of modern astronomy, the enormity of the Universe has instead given us the impression that we are just tiny particles, almost nothing at all. The Earth is nothing more than one of the planets in the suburbs of our galaxy that move around a small undistinguished star called the Sun. And we know that there are thousands of millions of galaxies like our own, and when our technological knowledge allows us to see further, that number may well increase many times over.

Geocentrism and heliocentrism: a crucial conflict

Humanity, in its efforts to understand the cosmos, began by forming models of a geocentric and heliocentric character. In the primitive cosmologies of ancient cultures, the Earth appeared as the solid centre of the Universe. It is, in fact, quite a reasonable outlook to adopt if your interest does not really lie in finding explanations for the motions of the heavenly bodies; it corresponds with intuition and is of an elegant simplicity. The Babylonians and the Greeks chose this option.

According to Plato, the Earth is a sphere in the centre of the Universe. The stars and the planets revolve around it, describing circles in this order from the inner to the outer: the Moon, Sun, Venus, Mercury, Mars, Jupiter and Saturn and then the stars. Aristotle offered a more elaborate system: the spherical Earth is situated in the centre of the Universe, and all the heavenly bodies are stuck on 56 concentric spherical surfaces around it, with several spheres for each planet.

Given the only observations that were available to mankind at that time, it is not surprising that the geocentric model would be the one they preferred. As we have seen, the Earth's rotation gives rise to a motion relative to the celestial sphere and all its stars. The shapes of the constellations, whose names we inherited from ancient Greece, do not change throughout the year. No kind of parallax with the stars can be seen as the great distances mean that the parallax is unimaginably slight.

The big problem crops up when attempts are made to explain the movements of the planets. These bodies appear as 'wandering stars' which travel around on the starry background even describing, in some cases, loops, advancing and retreating in an inexplicable fashion in comparison to the overall motion of all the other stars. The concentric models they used were also unable to provide an explanation for other visible phenomena, such as changes in brightness. This strange behaviour was the reason that many cultures saw these celestial bodies as a manifestation of the gods. And that's why their present-day names correspond to what at one time were Greek divinities and their Roman equivalents.

A time-delay photograph showing the apparent retrograde movement of Mars (source: Tunç Tezel, TWAN).

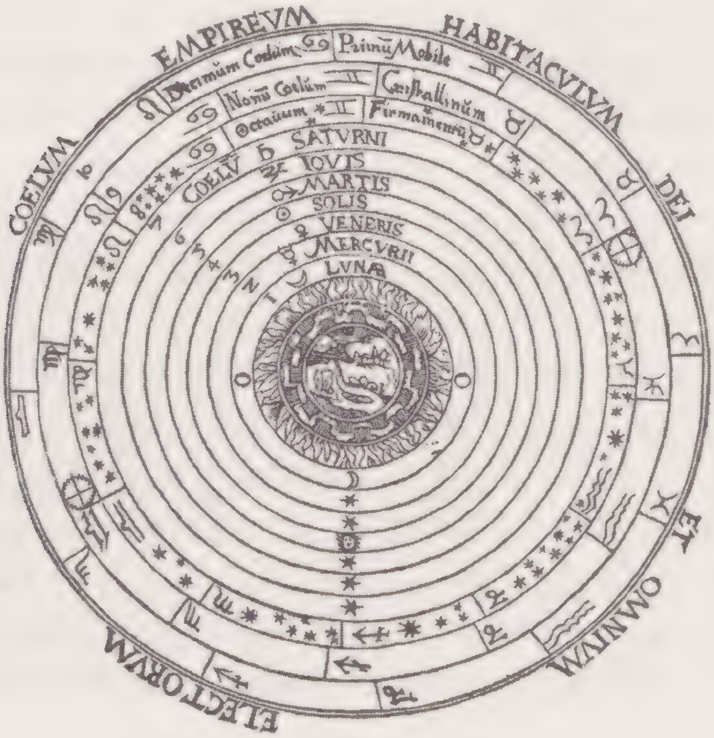
The Ptolemaic system and the Egyptians

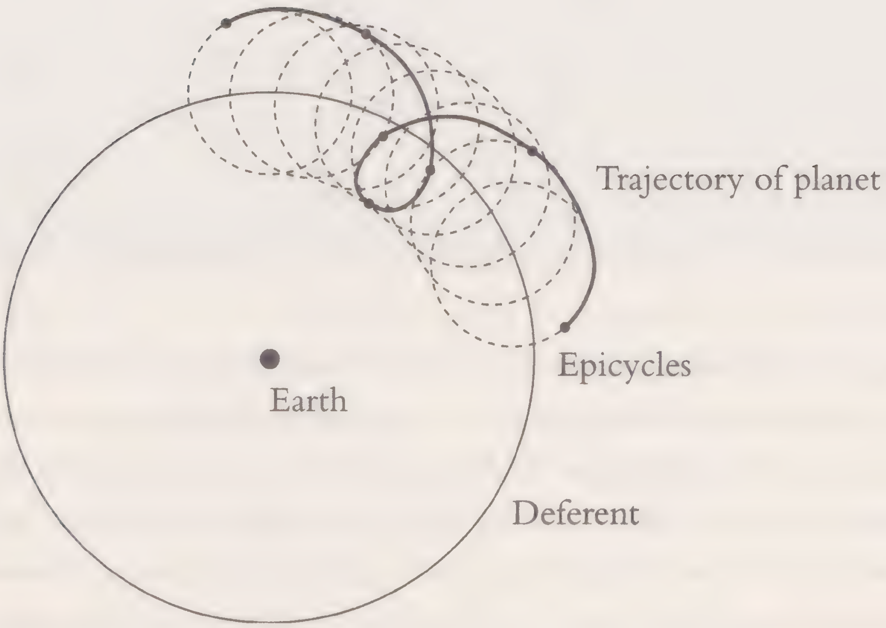
In the 2nd century AD, Claudius Ptolemy presented his cosmological model which was generally accepted by Christian and Islamic astronomers for the next millennium. In his great work *Almagest*, Ptolemy gathered together the work of preceding Greek astronomers and offered an explanation for the planets' strange behaviour. According to Ptolemy, each planet moves in accordance with the interrelation of several spheres.

The first of these is the 'deferent', and is centred on the Earth or slightly deviated from it, in a somewhat eccentric position. The other sphere, the one on which the planet actually moves is the 'epicycle' and it is centred on an arbitrary point on the deferent on which, curiously, there is no body. The combined motion of both spheres makes the planet move away from, and closer to, the Earth, reduce its speed until it stops and afterwards continue to advance in the opposite direction, which

thus explains the retrograde motion, which can be seen in many planets, with respect to the celestial sphere.

Ptolemy's order of the spheres from the Earth is the Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn and the fixed stars.

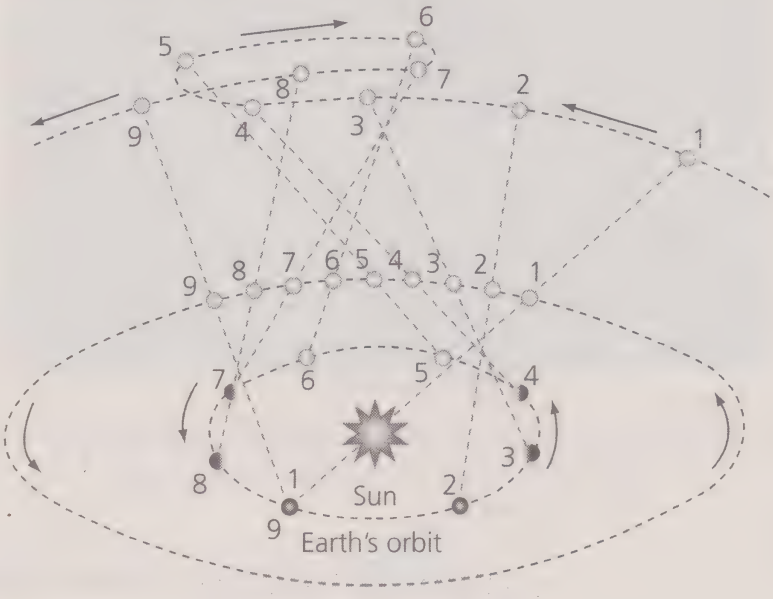




Top: Ptolemy's geocentric model. Above: the basic elements of Ptolemy's astronomy showing the path of a planet on a deferent-centred epicycle.

THE LOOPS OF MARS

The outer planets can be seen from Earth as wandering stars that advance and retreat as a result of relative movement. For example, in the diagram below, when the Earth moves through positions 1, 2, 3 and 4, we see the planet advance. When the Earth passes from position 5 to position 6, we see it recede, and then again advance from positions 7 to 8 and 9. Except for Mercury and Venus, this takes place with any of the outer planets. The loops followed by Mars are the easiest to detect as it is the fastest-moving of the outer planets and less time is needed for the phenomenon to be observed. It is possible to create an image showing the motion of Mars, like the one on page 36, simply by taking a set of photographs from the same spot (use a tripod!) and superimposing them on a computer.



Even with this sophisticated model, there were still some observations that could not be explained. For example, the loops that Mars follows are not always of the same size and, with the system Ptolemy established, they are impossible. To get over this drawback, the 'equant' was introduced, which is a point near to the centre of the planet's orbit from which the centre of its epicycle seems to move at the same speed, and so the planet moved at different speeds when the epicycle was at different positions in the deferent. This system was widely accepted as at that time it was able to explain the retrograde movements quite well. It was, in reality, very problematic,

as each planet had its epicycle revolving round a different deferent. What's more, it seems very strange that something should revolve around an empty point in space. Why this point and not another, as there was nothing at the other point, either? The fact is that with epicycles, almost any path can be generated. The secret is in considering enough epicycles and deferents to be able to explain the path of any planet. In the network of epicycle systems it is even possible to find examples that describe the figure of the cartoon character Homer Simpson!

Not all Greek thinkers accepted the geocentric model: some believed that Mercury and Venus moved in epicycles around the Sun, while still believing that the other planets did the same in respect to the Earth. But the most interesting model, from our point of view, is the one created by Aristarchus of Samos (c.310 BC–c.230 BC). In a book that has not survived through to our times, but which we know of from references to it by Plutarch and Archimedes, Aristarchus placed the Sun in the centre of the Universe, with the Earth and the other planets revolving around it. By using his ingenuity and diverse observations, Aristarchus determined that the Sun was much bigger than the Earth, which is why he concluded that the latter revolved around the former. His theory was not generally accepted; in fact it had no followers at all, which was a pity. If it had been the other way round, there is no doubt that astronomy would have made giant steps forward, but it was necessary to wait until the Copernican revolution in the 16th century to begin to speak of heliocentrism in a serious manner. In the Appendix there is a simplified explanation of Aristarchus's process for determining the relationships between distances and diameters of the Earth-Moon-Sun system which shows the Greek's great creativity and ingenuity in reaching his conclusions.

Heliocentrism: the simplest explanation

Technological advances often bring about a new realignment of the explanations accepted up to that moment by astronomers. Observations can turn out to be unwelcome, but salutary, shocks and prevent scientific knowledge from stagnating. Although these changes in concepts are not easily accepted by the community, a scientist should not be frightened of opening new channels. This is a road full of obstacles, where the mystery of how far one can get makes scientific thinking seem like a great adventure.

One of the obstacles that throughout history has appeared in the pathway of the scientific adventurer is religion and, for historical and geographical reasons as well as

those of dogma, especially the beliefs of the Catholic Church. This institution, which had adopted the role of custodian of all knowledge, did not welcome the changes to astronomical observation following the introduction of the telescope – in a similar sort of way to how a certain part of society seemed to feel offended when Pluto was 'downgraded' to the status of dwarf planet in 2006.

Nicolaus Copernicus (1473–1543) took 20–25 years to write *De Revolutionibus Orbium Cœlestium* (*On the Revolutions of the Heavenly Spheres*), in which, contrary to the official doctrine of the time, he claimed that the Earth and the other planets revolved in circles around the Sun, and that the centre of the Universe is a point near to it. Moving out from the Sun, are Mercury, Venus, the Earth with the Moon orbiting around it, Mars, Jupiter and Saturn. The stars are objects that are much further away than the Sun and which stay fixed without revolving in respect to the Sun. The retrograde motion of the planets is a consequence of the Earth's motion, and the Earth has, in addition to its translational motion, a rotational motion, with the inclination of its axis being the explanation of the seasons. It should be mentioned that the first copy of the published book reached Copernicus on the day he died, in 1543. Copernicus's book was revolutionary in the full sense of the word, and gave rise to great changes in science. It seems that the Polish astronomer feared the reaction of the Church – especially his uncle and patron, a powerful Polish cleric – and for years chose not to publish it.

A postage stamp commemorating the 500th anniversary of the birth of Nicolaus Copernicus. Along with an engraving of the astronomer, the stamp reproduces the front cover of De Revolutionibus *and the heliocentric model Copernicus proposed within the book's pages.*

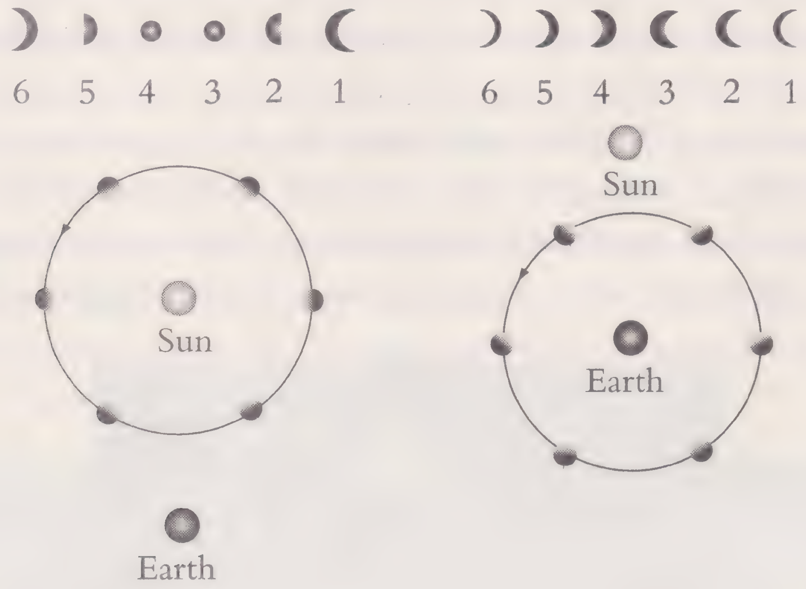
While working in the administration of his diocese, Copernicus spent long periods of time on patient and meticulous observation of the skies, and the information he gathered was later poured into *De Revolutionibus*. Copernicus himself carried out

a review of those issues that could not be adequately explained by the geocentric model: the lack of precision in predicting the annual movement of the Sun and the movements of the Moon; the explanations of the movements of the planets for which homocentric circles were used in some cases and at other times eccentric circles, along with the equant and the epicycles... it all added up to a Universe lacking unity. The ideas that this genius put forward, on the other hand, showed unity of motion and a rational system of circles, with no equants. He overcame the parallax problem by speculating that it was due to the huge distance separating us from the stars, and he determined the Earth's motion as being the ultimate reason for the movements on the celestial sphere. Neither were the Sun's movements due to the star itself, Copernicus argued, but to an Earth that was spinning in a manner analogous to the rest of the planets and, in the same way, the direct and retrograde motion of the planets were not due to the planets themselves, but to relative motion with respect to the Earth.

The protagonist of the other great revolution was the Italian, Galileo Galilei (1564–1642) who, in 1609, first used a telescope to observe the sky. Up to that time, only the military appeared to have taken any interest in telescopes, but from then onwards, telescopes took on a crucial role in the development of astronomy.

A portrait of Galileo from 1636, by the Flemish artist Justus Sustermans.

Among Galileo's numerous and extraordinary observations, that of Jupiter's satellites, of which he observed four on the night of 7 January 1610, was a particularly hard blow to geocentrism. It was clear that not all heavenly bodies revolved around the Earth: there were at least four that did not. But Galileo also observed that Venus's phases were incompatible with the Ptolemaic model, in which Venus is located on a sphere between Mercury and the Sun in such a way that the phases should always have diameters that are more or less analogous. Venus, however, is seen in full phase when it is small, and when it is in waxing or waning phase it appears much larger, as can be seen in the diagram below. Even Galileo did not dare publish his observations of Venus until 1623. With this new proof, the Ptolemaic system was mortally wounded.



On the left, the phases of Mercury or Venus according to a heliocentric model. The diameter of the phase has a different size depending on the distance from the Earth. On the right, the same phases according to a geocentric model. The phases always have approximately the same diameter, and the epicycles can only explain a small change in size, which does not tally with real observations.

In a desperate attempt to maintain the rejection of Copernicus's radical heliocentrism, the Jesuits of the Roman College searched for an alternative in the system proposed by the Danish astronomer Tycho Brahe (1546–1601), who had seemed destined for oblivion up to that time. Brahe believed that the Earth was unmoving, as in his observations he was unable to perceive the parallax of the stars, and that the Sun and the Moon revolved around the Earth, while Mercury, Venus, Mars, Jupiter and Saturn revolved around the Sun in circular orbits.

According to Tycho Brahe, the Solar System was intermediate between Copernicus's heliocentric and Ptolemy's geocentric. As mentioned above, Brahe thought that the Earth was unmoving, as he could not see the parallax of the stars. So, the Sun and the Moon revolved around the Earth while Mercury, Venus, Mars, Jupiter and Saturn revolved around the Sun, all of them moving in circular orbits.

The system proposed by Tycho Brahe, halfway between geocentrism and heliocentrism.

In his monumental *Sidereus Nuncius* (*Sidereal Messenger*) of 1609, Galileo rejected the idea that the heavens were perfect and unchanging, as Aristotle and Ptolemy claimed, because there are mountains on the Moon, as his observations made clear (he had even calculated their height), and it was therefore not smooth and unchanging. He also stated that the number of visible stars was multiplied with the telescope, but their size did not increase as it did with the planets or the Moon. This proved Copernicus' theory that there was a large gap between Saturn and the fixed stars and it explained the absence of the parallax which, according to the arguments of the geocentrists, should appear as a result of changes in the Earth's position in its orbit. So it was clear that the lack of parallax was not due to the fact that the Earth did not move, but because the stars were very far away.

More evidence of the imperfections of celestial objects and their changeability was provided by Galileo in 1612 when he published his observations on sunspots. Furthermore, as a result of his studies of sunspots, he proved that the Sun rotated on its own axis, which to a certain degree could suggest that the Earth perhaps did the same.

In his defence of the Copernican system, Galileo always provided evidence from observation of reality by using scientific methods for the first time. However, as his evidence apparently contradicted common sense – the Earth can't be seen to move, but the Sun can! – it clashed with the ideas of intellectuals and theologians clinging to a literal reading of the Holy Scriptures.

The Galileo affair

Galileo's arguments seemed irrefutable and convincing, but the supporters of the geocentric theory would not give way, though they used arguments which, from our modern viewpoint, may seem rather surprising. For example, Martin Horky published a paper in 1610 in which, in addition to personal insults, he denied the existence of the Galilean satellites: "Astrologists made their horoscopes by taking into account everything that moves in the heavens. Therefore, the Medici Stars (the satellites discovered by Galileo) serve no purpose whatsoever and, if God does not create useless things, those stars cannot exist."

Other similar attacks were launched, but the situation worsened in 1615 when Galileo wrote a letter to Christina of Lorraine in which he set out his arguments on the possible discrepancy between his discoveries and the Holy Scriptures. In the letter, he claimed that in the Bible: "The intention of the Holy Ghost is to teach us how one goes to heaven, not how heaven goes." This letter played a key role in the subsequent proceedings. Galileo went to Rome to defend himself before the Holy See, but he lacked irrefutable evidence that the Earth rotated. What would he have given for Foucault's pendulum! He attempted to prove the rotation with his theory on tides, the only one of his theories that he got wrong, while the Jesuits maintained, correctly, that tides were caused by the attraction of the Moon.

Finally, in 1616, the Copernican theory was condemned as being "foolish and absurd in philosophy, and formally heretical", and Galileo was requested to set out his work as a hypothesis and not as irrefutable truth. This situation, together with his poor health, deeply affected the Italian, and drastically reduced the amount of work he carried out. A few years later, however, he started up again and set out to

study Jupiter's satellites and to tabulate its movements. In 1632, Galileo published *Dialogo Sopra i Due Massimi Sistemi del Mondo* (*Dialogue Concerning the Two Chief World Systems*), in which he presented the Aristotelian system and the Copernican by following an original idea of Pope Urban VIII's, under the protection of the Grand Duke of Tuscany, Ferdinand II of Medici. Although Urban VIII's idea was to give an impartial presentation of both systems, Galileo blatantly biased the book in favour of the Copernican system. The work caused a tremendous scandal and an authentic revolution in the scientific world of the time, and sparked a furious response from the papacy against Galileo. Malicious gossip, possibly put about by Galileo's enemies, claimed that one of the three characters in the dialogue, Simplicius ('the simple one'), was based on Pope Urban VIII himself.

It has to be acknowledged that in the *Dialogue*, Galileo provided two unconvincing experimental proofs with the aim of demolishing Tycho Brahe's system, which was the one defended by the Jesuits. The proofs were based on his theory on the tides, which was erroneous, and on the rotation of sunspots, which was also compatible with Brahe's system. However, this publication provided sufficient reason for the Inquisition to again intervene, as Galileo was only permitted to present his theories as mere hypotheses, and not to offer evidence to back them. More than that, it turned out that Galileo, instead of using Latin, had written his *Dialogue* in everyday Italian, making it more accessible and therefore annoying the Church even more.

Galileo Before the Holy Office in the Vatican, 1632,
painted in 1847 by the French artist Joseph-Nicolas Robert-Fleury.

The Inquisition charged Galileo with breaching the 1616 ban, and he was summoned towards the end of 1632 to appear in Rome "willingly or otherwise". The astronomer was initially treated fairly but in the end was urged to confess and threatened with torture. On 30 April, Galileo admitted the charge and was sentenced to life imprisonment with the option to recant – which he did. The Pope subsequently reduced Galileo's sentence to house arrest for life.

In 1638, Galileo went blind due to so much observation of the Sun and, to continue working, he had to rely on help from his followers. Thanks to occasional visits from colleagues, his works were smuggled across the border and published in Leiden and Paris. Galileo died in 1642 and was buried in Florence. His works, and in particular the *Dialogue*, were decisive in introducing the essentials of the scientific method and in promoting currents of rational thinking. This prepared the way for the separation of science and theology.

THE TRAGEDY OF THE MOON

Some years ago I had the pleasure of reading a book of short stories by Isaac Asimov. One of them is entitled 'The Tragedy of the Moon' and it offers a very interesting hypothesis. It basically consists of imagining what the history of astronomy would have been like if the Moon had turned out to be not a satellite of the Earth but, instead, a satellite of Venus; Asimov even gives it a name: Cupid. This satellite, Cupid, would have the same characteristics as our Moon, but would revolve around Venus. The fact that we have the Moon as a satellite of the Earth made many ancient astronomers believe in geocentricism, since the Moon was seen to revolve around the Earth, and the Sun seemed to do the same. The outer planets showed strange behaviour, but we have to bear in mind that they did not show any clear signs pointing towards heliocentrism. The inner planets, Venus and Mercury, could have given more clues. But Mercury, being near the Sun, is difficult to observe, and Venus, which at its maximum rises only to about 47° (little more than two hands if we use our hand to measure angular distances), was observed as the morning star or the evening star – the ancients did not identify these 'two' as being one and the same planet, so they did not realise that it was just one planet revolving around the Sun. If all their observations were basically of motions that appeared to be geocentric, there was no reason for it to occur to them that other bodies, apart from the Moon, revolved around any body other than the Earth. But let's see what would have happened if the Moon, instead of being our satellite, had belonged to Venus. Let's work out what

In 1965, the Second Vatican Council acknowledged and apologised for the Church's improper interventions in science and cited the Galileo case: "We cannot but deplore certain attitudes which have existed among Christians themselves, insufficiently attentive to the legitimate autonomy of science." In 1992, John Paul II paid homage to Galileo in his speech at the Pontifical Academy of Sciences and acknowledged the mistakes committed by theologians in the 17th century: "The greatness of Galileo is well known; he was to suffer greatly – we cannot hide that fact – due to the actions of men and the Church." He apologised and proposed a commission be formed to carry out a full rehabilitation of the figure of Galileo.

Encouraged by the International Astronomical Union, UNESCO declared 2009 to be the International Year of Astronomy, and the Holy See held an international congress on Galileo in an attempt to mend bridges with the scientific world.

the sky would have been like, as seen from the Earth. Firstly, without the light pollution from moonlight, the observation of the skies would have been improved considerably. For many nights and early mornings, the brightest object in the sky would have been Venus. This Venus, together with its satellite Cupid, would have presented variations in luminosity, as it does now on account of its phases. Depending on the relative position of the Earth, Sun and Venus, we would see Cupid with greater or lesser luminosity. Its maximum value of brightness would be similar to that of the planet Saturn, or the star Arcturus. The extension of Cupid's orbit in respect to Venus as seen from the Earth would be 0.6°, that is, a little larger than the diameter of the Sun. So, for us, the brightest object in the sky would have been a satellite which was in plain sight and would at times have been seen as distant from the body of the planet as the diameter of the Sun. More than that, this phenomenon would have made it easy to see that the morning and evening stars were exactly the same object. It would be very clear, too, that Venus revolves around the Sun. In this way, the ancient astronomers would have had sufficient evidence of objects revolving around different bodies for the hypothesis of geocentric cosmologies to be inconceivable from the beginning. It is very likely that humankind would not have had to wait until the 16th century and Copernicus before making its first heliocentric studies. Looked at from this perspective, humankind's scientific development truly has been affected by 'The Tragedy of the Moon' orbiting the Earth.

The Kepler solids

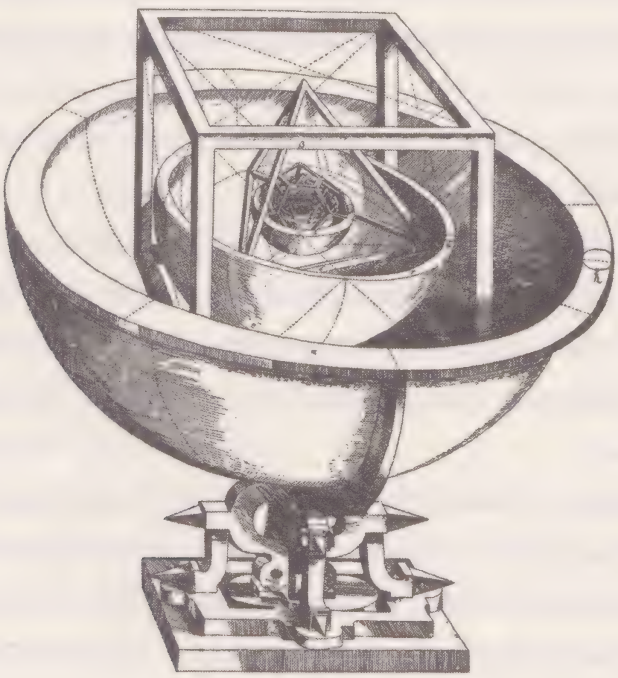
Tycho Brahe is considered to have been the best observer of the period prior to the use of the telescope. He built Uranienborg ('Castle of Urania'), a palace on the Danish island of Hven which he made into an observatory and research centre. He designed his own instruments and was able to determine the position of the stars and the planets with much more accuracy than was possible for other observers of that era. He compiled a star catalogue containing more than 1,000 stars whose positions were determined to within half an arc-minute. Night after night, Brahe made systematic observations with the highest accuracy possible. It was these observations that allowed Johannes Kepler to deduce his famous laws on orbits; in particular, the measurements of the movements of Mars played an essential role in discovering them.

In October 1600, Tycho Brahe, having read some of Kepler's works, invited Kepler to meet with him in Prague. Their relationship, however, was not easy and was riven by distrust, possibly due to their strong, rather arrogant characters. After Brahe's death, Kepler took his place as Rudolph II's imperial mathematician and at times he worked as the astrological adviser.

Tycho Brahe (above left), the Danish nobleman and astronomer. In this engraving by Johann-Leonhard Appold, Brahe is shown with a prosthetic nose, a consequence of having been injured in a duel by sword. Kepler (above right), of German origin, was tenacious and deeply religious, character traits quite different from Brahe's.

Kepler devoted most of his life to understanding the motion of the planets. It turned out that, by coincidence, the number of known planets in his epoch was one more than the number of perfect polyhedra. Euclid had proved that there existed five regular polyhedra, each of them inscribable in a sphere and circumscribable by a concentric sphere. They are what are known as the Platonic solids – polyhedra whose sides are regular convex polygons – the tetrahedron, the cube, the octahedron, the icosahedron and the dodecahedron. In Kepler's view, it could not be a coincidence that the number of spaces between the planets was five. Initially, he thought that the motion of the planets had to verify the Pythagorean laws of harmony, the so-called music of the celestial spheres, which we shall look at in more detail later. But, as Kepler believed in the heliocentric model, he attempted to prove that the distances from the planets to the Sun corresponded with spheres situated inside perfect polyhedra, and so on. In the most interior sphere was Mercury, and the other five planets (Venus, the Earth, Mars, Jupiter and Saturn) would be situated in their respective spheres in the interior of each of the five Platonic solids.

After years with no results and in view of the data, particularly those relative to the retrograde motion of Mars, Kepler was forced to acknowledge that the motion of the planets could not be explained by his model of perfect polyhedra and the harmony of the spheres. He then tried with all kinds of possible combinations of



A Platonic-solid model of the Solar System presented by Kepler in his work Mysterium Cosmographicum *(1596).*

circles. By now getting desperate, Kepler tried using ellipses. He was deeply religious and he found it difficult to believe that God would have used ellipses to plot the movements of the planets: "Why ellipses when there are circles?" he asked. But, with ellipses, he had more luck and was able to formulate his three famous laws explaining the motion of the planets.

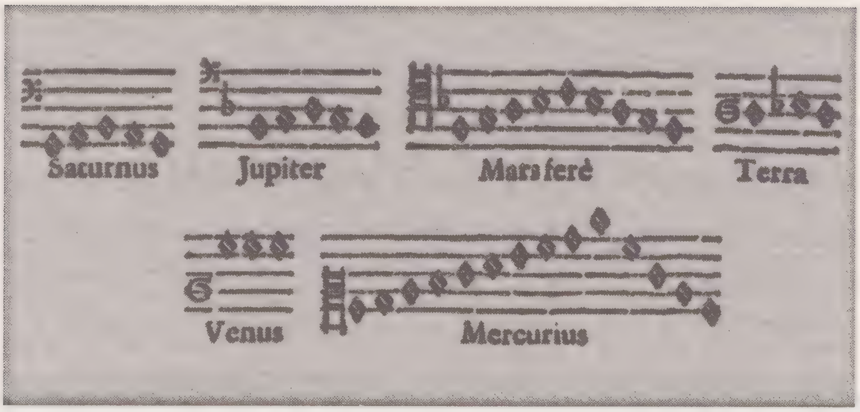
It is worthy of note that Kepler prioritised data gathered from observation in preference to theory, making him a truly modern astronomer. Basically, Kepler was right; nature describes simple figures. In his general laws of relativity, Einstein proved that in the tetradimensional geometry in space-time heavenly bodies follow straight lines. In other words, a line that is even simpler than a circumference. Confirmation of Kepler's laws arrived with the accurate prediction of the transit of Venus in 1631.

The music of the spheres according to NASA

For Pythagoreans, the distances between the planets' spheres had the same proportions as those between sounds on the musical scale considered 'harmonic', or consonant. Each sphere produced a sound as does a projectile cutting through the air, being higher or lower pitched depending on how near they were. The sounds made by the spheres combined to produce what was called 'music of the spheres'. Plato, who arrived on the scene much later, saw the world as a great animal equipped with its own soul, and in his dialogues claimed that the soul of the world had been created in accordance with the musical proportions discovered by Pythagoras. In his acclaimed dialogue, *Republic,* Plato described the planets as moving in a series of circles which revolved in concentric orbits.

Kepler studied Plato's myths and meticulously scrutinised the motion of the planets with the aim of finding the relationship between their motion and musical theory. He determined that a celestial body emitted a sound which is higher-pitched the faster it moves, prompting him to introduce well-defined musical intervals linked to the different planets.

In his book *Harmonices Mundi* (*The Harmony of the World*), Kepler claimed that each planet's angular velocity produced consonant sounds. He even wrote six melodies, each of them corresponding to one of the planets known at that time. Put together, these melodies produced four different chords, one of them the chord produced at the time of creation, while another was the chord that would sound when the Universe came to an end.



Kepler's planetary melodies.

Centuries later, Newtonian mechanics led to the discovery of Neptune through mathematical calculation, which strengthened the idea that the Universe was harmonic. Einstein, for example, was convinced of the harmony of the Universe. The new language of astrophysics with its terms such as spectra, frequencies, resonance, vibrations and harmonic analysis states that a time-varying signal can be described by a composition of trigonometric functions.

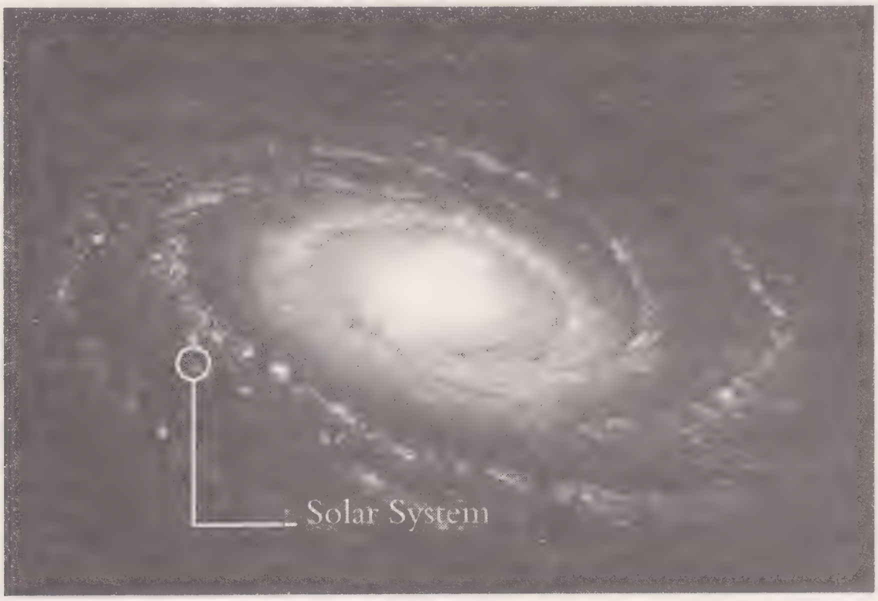
One of the latest theories in physics describes elementary particles not as corpuscles but as vibrations of minute strings, considered to be geometric entities of one dimension. Their vibrations blend into special mathematical symmetries which represent a prolongation of the Pythagorean view of the world and a return to – with a modern-day viewpoint – the ancient belief in the music of the spheres.

In April 1998 NASA's TRACE (Transition Region and Coronal Explorer) satellite found the first evidence of music created in a heavenly body, just as had first been supposed by Pythagoras, and later by Kepler. The purpose behind TRACE was to study the Sun's turbulent upper atmosphere, the crown, where solar storms are unleashed. Scientists at the Southwest Research Institute in Texas discovered that the star's atmosphere, packed with ultrasound in the form of waves, really does 'sound', just as the Pythagoreans had predicted.

According to this discovery, the traditional music of the spheres really consists of a solar ultrasound which plays a music score composed of waves pitched 300 times lower than the frequency of the deepest waves that can be heard by a human ear, with a frequency of 100 mHz (millihertz). Humans cannot hear sounds of a frequency lower than 16 Hz (infrasonic sounds), or over 20 kHz (ultrasonic or supersonic sounds). The music produced by the Sun is not audible to us.

What is a planet? And what isn't a planet?

At its general assembly in Prague in 2006, the International Astronomical Union (IAU) declared that Pluto was no longer a planet. The author of this book admits that she voted for the eviction of 'poor' Pluto. Let's take a look at the reasons behind this drastic decision, which was taken, let's not forget, by thousands of professional astronomers.



The position of the Solar System within the Milky Way, our galaxy.

The Solar System is a planetary system situated in one of the branches of the Milky Way galaxy. We are really situated in a 'suburban' part of the galaxy, some 8.5 kiloparsecs, or 28,000 light years, from its centre. The Solar System has just one central star, the Sun, with diverse bodies moving around it, practically on the same plane (the ecliptic plane). All of these bodies follow elliptic orbits, spinning anticlockwise from the perspective of the solar north pole.

The planets are bodies that describe paths (orbits) while revolving around the Sun. They are of sufficient mass for their gravity to overcome the interior forces as a rigid body (for which reason their shape is practically spherical), and to clear their orbits' surroundings of smaller objects (planetesimals) by capturing them. The Solar System has eight planets which are divided into inner ones, also called terrestrial or telluric planets (Mercury, Venus, Earth and Mars) and outer, or gigantic planets (Jupiter, Saturn, Uranus and Neptune); all the outer planets have rings around them.

The dwarf planets (a new category of body orbiting the Sun, which was defined in August 2006) are objects whose mass is large enough to enable them to have a

spherical shape, but not enough to have attracted or rejected all the small bodies around them; Pluto, Ceres, Eris, for example, are in this group, along with many others. The asteroids, too, are minor bodies whose small mass has not enabled them to achieve a spherical shape. They mostly accumulate either in the belt of asteroids situated between Mars and Jupiter, or in the Kuiper belt, beyond Neptune.

There are also satellites, major bodies that orbit some planets in elliptic orbits, and finally, the comets: small frozen objects from the Oort Cloud which describe elliptic, hyperbolic, or parabolic orbits. Additionally, it must be added that in the space around the Sun, there is interplanetary dust composed of microscopic solid particles and tenuous gas with charged particles forming a plasma expelled by the Sun in what is called solar wind. The boundary of the Solar System is calculated at 100 AU.

Planet	Diameter (km)	Distance to the Sun (km)	Distance to the Sun (AU)	Inclination (in respect to the ecliptic)
Sun	1,392,000			
Mercury	4,878	$57.9 \cdot 10^6$	0.38	7.0°
Venus	12,180	$108.3 \cdot 10^6$	0.72	3.4°
Earth	12,756	$149.7 \cdot 10^6$	1.00	0.0°
Mars	6,760	$228.1 \cdot 10^6$	1.52	1.9°
Jupiter	142,800	$778.7 \cdot 10^6$	5.20	1.3°
Saturn	120,000	$1.430,1 \cdot 10^6$	9.55	2.5°
Uranus	50,000	$2,876.5 \cdot 10^6$	19.22	0.8°
Neptune	45,000	$4,506.6 \cdot 10^6$	30.06	1.8°

Details of the planets in the Solar System.

The case of Pluto

Pluto was discovered in the 20th century by photographic techniques. It was an amateur US astronomer, Clyde William Tombaugh, who discovered Pluto while carrying out a systematic search using the 'blink' technique in 1930. This technique consists of taking two photos of the same region with an interval of time sufficient for an object which has moved to seem to 'jump' when the two images are compared. This was how Tombaugh managed to discover an object beyond Neptune which was moving against the starry background (after comparing more than 15 million stars!).

Pluto, its satellite Charon and two smaller moons to their right – Nix (above) and Hydra.

Shortly after being discovered, Pluto was classed as the ninth planet by the IAU. But Pluto had certain features that made it stand out when compared to the other eight. All of them orbit approximately on the same plane of the ecliptic, but Pluto orbits with a degree of inclination of 17.2° in respect to the ecliptic, as do a large number of the objects that form the Kuiper belt.

At the turn of the 21st century, three bodies of a similar size to Pluto were discovered in the same zone. In August 2006, the IAU had to choose between increasing the number of planets in the Solar System from 9 to 12 – with the possibility that this number would increase enormously in the future on account of technological advances – and reducing the number to eight. That is how and why the polemical Pluto, on account of its small size and dynamic evolution, was definitively excluded from the catalogue of planets and became 'just' a dwarf planet, like Ceres, Makemake, Haumea and Eris. There are those who view these changes as a lack of scrupulousness on the part of astronomers but here, on these pages, we want to emphasise the scientific rigour underlying Pluto's new status. These changes are simply the result of advances made in the observation of the Solar System. Nowadays, many more objects are known to exist in the Solar System than were known at the beginning of the 20th century. If now, on account of that new knowledge, corrections have to be made to data that were previously acknowledged and accepted, that is the sign of a serious and rigorous way of operating. Every astronomer has to be prepared to modify their working hypotheses in line with new discoveries.

SCALE MODELS OF THE SOLAR SYSTEM

As mentioned previously, it is not easy to get an idea of the size of the Solar System. To comprehend its size, we propose constructing a simple scale model that will give a clearer idea of its true dimensions. The best way is to visualise a model representing the bodies to scale and at their relative distances. The problem is that it is not a simple task to find a scale that allows the planets to be represented by objects that are not too small to see, and the distances between them overwhelmingly great.

Let's take a model made up of balls of various diameters. At one end of a square in a town or village, or at one end of a park, we place a basketball of about 25cm diameter. This ball represents the Sun. Mercury will be a pinhead (1mm in diameter) situated 10 metres from the Sun. Another pinhead, or perhaps the head of a needle, somewhat larger this time (2mm in diameter), will be Venus, located 19 metres from the Sun. The Earth is the head of another pin like the previous one (2mm) 27 metres from the Sun. Mars is another pinhead, a smaller one (1mm), 41 metres from the Sun. The next planets would have to be placed even further away. A table-tennis ball (25mm in diameter) would be Jupiter, 140 metres from the Sun. Saturn would be another ball 20mm in diameter and 250 metres from the Sun. A marble (10mm in diameter) would represent Uranus, and would be placed 500 metres from the Sun. And finally, another marble (10mm) 800 metres away, would be Neptune. It can clearly be seen that the Solar System is quite 'empty', particularly when the planets are not aligned as they are in the model set out above, but move around on their different orbits and are therefore much further away from each other than in this example. Some cities and towns have chosen to display the Solar System in model form in a public place. For instance, in France, the city of Metz has laid out a Solar System in the city streets and squares showing the different planets and there are information panels for passers-by keen to know more.

Can two planets collide?

This is a very common question that children ask when someone speaks to them about the Solar System. We have seen how the Solar System is very 'empty'. For two planets to move out of their orbits, there would have to be some exterior body causing a large gravitational upset. This body would have to be huge to affect a planet in such a way that it leaves its orbit and that possibility is very remote. Much more likely are impacts on planets from the Solar System's minor bodies,

such as asteroids and comets. Barringer Crater in Arizona, the best known of all terrestrial craters, is a consequence of the impact of a meteorite on the Earth. The Moon is pitted by impact craters due to the fact that its minimal atmosphere does not behave like the Earth's atmosphere, which is able to burn up most objects before they reach the surface. But such impacts also happen on other planets: in July 1994, 21 fragments of the comet Shoemaker-Levy 9 penetrated the atmosphere of the planet Jupiter and the impact sites could be observed from Earth. It is evident that at the time of the Solar System's formation, such impacts must have been much more frequent.

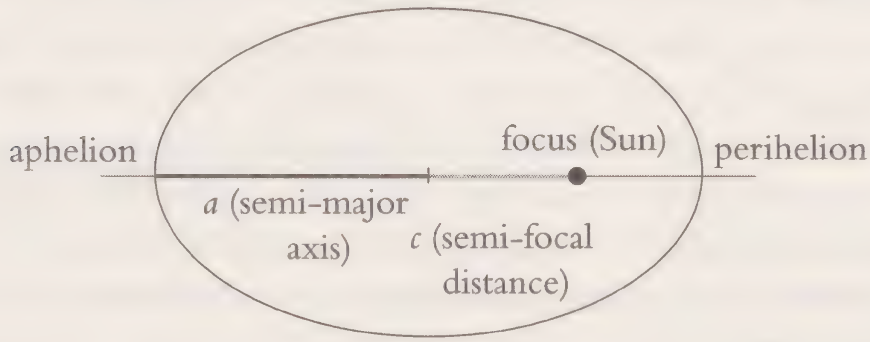
The 'barcode' of the planets

Celestial mechanics is able to describe the trajectories of the planets and to predict all kinds of transits and relative positions between them. This is, in fact, one of the characteristics of astronomy as a science – the ability to predict the position of the heavenly bodies and the existence and visibility of certain phenomena. To be able to do this, we use what are called orbital elements of the Solar System bodies. We shall describe them here specifically for the planets. These elements are something akin to the 'barcodes' of the planets, as they give all the information required to be able to know their orbits and to be able to calculate them with complete accuracy.

Kepler determined the orbital elements and introduced them so as to facilitate the study of the motion of the planets around the Sun. The calculation methods he used were later put to good use by Newton, Gauss, Laplace and Olbers. Though here we shall give the details for the planets, in actual fact the orbital elements establish the orbits of all the bodies, all those objects that have mass, whether planets, asteroids, comets or artificial satellites.

The elements of a planet's orbit are a set of six values which enable a perfect definition of its orbit around the Sun, situated on the focus of the ellipse. The first three are known as the Euler angles and are used for defining the orientation of the orbit in space. The other three give the shape of the orbit and the position of the planet within it. These six values are the longitude of the ascending node Ω, the inclination of the orbit i, the argument of the perihelion ω, the semi-major axis of the orbit a, the eccentricity e and the epoch of passing along the perihelion Mo. Let's first look at more details on the last three, which define the type of ellipse (size and shape) and the position of the planet on it.

The semi-major axis a of the orbit corresponds to half of the major axis of an ellipse. The point of intersection of the semi-major axis with the orbit which is nearest the Sun is called the perihelion, and the most distant intersection point is the aphelion, as shown in the diagram below. So, the distance between the perihelion and the aphelion is double the semi-major axis of the ellipse. Another way of getting the size of the ellipse is to replace the semi-major axis by the period of time that the planet needs to make a complete orbit. Either of the two systems will obviously provide us with an idea of the size.



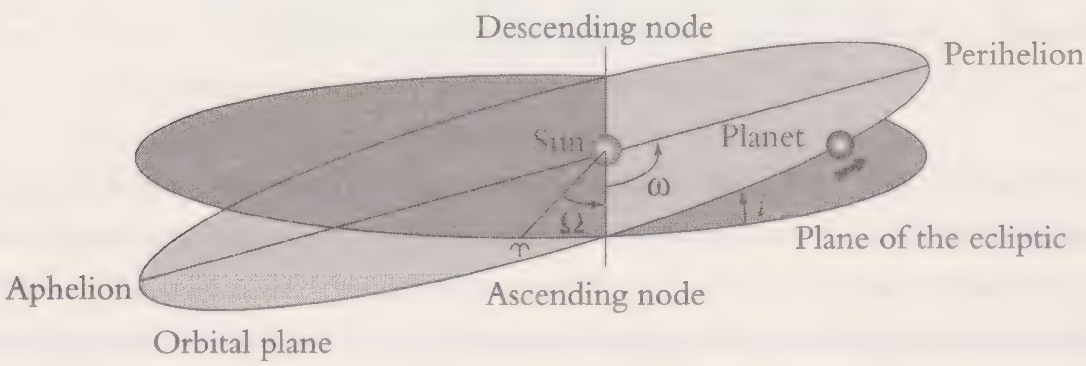
An elliptical orbit with the semi-major axis, the semi-focal distance, the aphelion and the perihelion, with the Sun being the body situated on the focus. The eccentricity e *is the product of the formula* $e = c/a$.

The eccentricity e of the ellipse serves to show how elongated it is. It is defined as the semi-distance c between the two foci of the ellipse divided by the semi-major axis a, that is $e = c/a$. If the orbit is closed, ie has a circumference, as the two foci meet at one point called the centre, the distance between the foci is zero, and consequently the eccentricity is zero. If an orbit has a very small eccentricity, practically zero, it has a shape which is close to being a circle, which is what happens, as we shall see later, in most of the planets' orbits. The eccentricity of an ellipse is always less than 1 because the semi-focal distance is, in all cases, less than the semi-major axis.

When the eccentricity is equal to 1, it is a parabola, an open curve, and it does not correspond with the case of the planetary orbits. There can be cases of comets with orbits showing eccentricity of even above 1, that is, when they follow hyperbolic orbits. In such cases the comets only come close to the Sun on one occasion; they approach the perihelion and afterwards do not reappear. This makes them the most spectacular, much more so than the elliptical comets which periodically approach

the Sun and gradually lose part of their mass on each run until they finally disintegrate into small pieces. The time of the perihelion passage, when the planet is at the point nearest the Sun, enables us to localise its position in its orbit.

Let's now look at the three other elements that enable us to indicate the position and orientation of the orbit in respect to the ecliptic. The inclination i gives the angle from the plane of the ecliptic to the plane of the respective planet's orbit. The line of intersection between those two planes is called the nodal line. In the diagram below, in which the ecliptic plane is shown as being horizontal, when the planet goes round its orbit it will have an ascending node (where the planet ascends from below the ecliptic) and afterwards a descending node (where it does the opposite). To fully determine the position of the orbit with respect to the ecliptic, another angle is needed, ie the longitude of the ascending node (Ω), the angle measured anticlockwise from the point of Aries (γ) to the ascending node. Finally, to orient the orbit on its plane, the third Euler angle is used, that is, the argument of the perihelion ω, the angle anticlockwise from the ascending node to the perihelion.



Ellipse of an orbit where we can see the inclination of the orbit i, *the longitude of the ascending node* Ω *and the argument of the perihelion* ω.

These orbital elements are also used to determine the orbits of other bodies in the Solar System and to plot the paths of artificial satellites. The six elements mentioned above are involved in relation to the two-body problem, ie two bodies with no external perturbations. A real perturbed trajectory is shown as an instantaneous sequence of cones that shares one of its foci. These orbital elements are then called osculatrices, and the trajectory is always tangential to this sequence of cones.

The orbital elements of real objects tend to change over time due, mainly, to the gravitational force of other bodies nearby in the Solar System. The case of Neptune

has already been mentioned; this was detected due to unexplained alterations in the orbit of Uranus. In another example, that of comets, the emission of gases, the pressure of radiation or electromagnetic forces can vary their orbits. On the other hand, changes in the orbital elements of artificial satellites may be a consequence of the Earth's lack of perfect sphericity, or contact with remnants of the terrestrial atmosphere. There are lots of computer programs for following artificial satellites' orbits round the Earth, but to be able to get accurate results it is necessary to constantly enter the values of the perturbed orbital elements; if the programs are not updated with the latest orbital elements, the results can be very inaccurate after just a month.

What are exoplanets? Do they exist?

The Solar System is believed to have been formed some 4,500 million years ago from a cloud of gas and interstellar dust which made up the central star and a circumstellar disc. Through the gathering together of small particles in the disc, other larger ones would have been formed, then planetesimals, later protoplanets, and finally the present-day planets. There is nothing to make us believe that it was a unique process, but rather that the same thing has happened on many occasions and has created other planetary systems which, as a central body, have stars other than the Sun.

The number of planets discovered outside the Solar System runs into several hundreds, and the majority of them are distributed in planetary systems with more than one planet. They are known as exoplanets and, with few exceptions, are all large – bigger, even, than Jupiter, which is the largest planet in the Solar System. That is why the masses of extrasolar planets are often compared with that of Jupiter ($1.9 \cdot 10^{27}$ kg). Only some of them are of a size similar to that of the Earth, though that may be due to our own technological limitations.

The exoplanets' nomenclature is simple – a lower case letter starting from 'b' is placed after the name of the star for the first planet found in the system (for example, 51 Pegasi b). The next planet is labelled with the next letter in the alphabet: c, d, e, f… (51 Pegasi c, 51 Pegasi d, 51 Pegasi e and 51 Pegasi f…).

In the table on page 60, there are some exoplanets that are very near the central star (Gliese 876 b, c, d, with orbits nearer to the star than Mercury is to the Sun). Others have more distant planets (HD 8799 has a planetary system with three planets situated more or less at the same distance as Neptune is from the Sun).

The first directly observed exoplanet 2M1207 b. It has a mass 3.3 times that of Jupiter and orbits at 55 AU from its central star, a brown dwarf. There is a disc of dust around the central star, which indicates the formation of other planets. (photo, ESO).

Name of planet	Mean distance (AU)	Orbital period (days)	Minimum mass*, (x masses of Jupiter)	Year covered.	Approximate diameter** (km)
Gliese 876 b	0.208	61.94	1.935	2000	
Gliese 876 c	0.13	30.1	0.56	2000	
Gliese 876 d	0.021	1,938	0.018	2005	
υ And b	0.059	4.617	0.69	1996	~Jupiter 140,000
υ And c	0.83	241.52	1.98	1999	~Jupiter 345,000
υ And d	2.51	1,274.6	3.95	1999	~Jupiter 297,000
Gliese 581 b	0.041	5.368	0.049	2005	~Earth 20,000
Gliese 581 c	0.073	12.932	0.016	2007	~Earth 14,000
Gliese 581 d	0.25	83.6	0.024	2007	~Earth 14,000
HD 8799 b	24	36,500	10	2008	
HD 8799 c	38	69,000	10	2008	
HD 8799 d	68	170,000	7	2008	

The table shows some extrasolar systems with multiple planets. Data taken from the Extra-solar Planets Catalogue (2) (except the last column).

** The radial velocity method used to detect them, it only gives the minimum mass of the planet.*

*** The diameter has been calculated from the cube root of the mass ratio times the diameter of the solar planet. (Wikipedia)*

We now know that there are exoplanets at different types of stars. In 1992, radio astronomers announced the discovery of planets around the pulsar PSR1257+12. This detection is believed to have been the discovery of the first exoplanet. In 1995, the first detection of exoplanets around a star of the G-51 Pegasi type was announced, and since then discoveries have been made of exoplanets around a red dwarf (Gliese 876 in 1998), a giant star (Edasich in 2001), a brown dwarf (2M1207 in 2004), a star of type K (HD40307 in 2008) and a star of class A (Fomalhaut in 2008 and 2012).

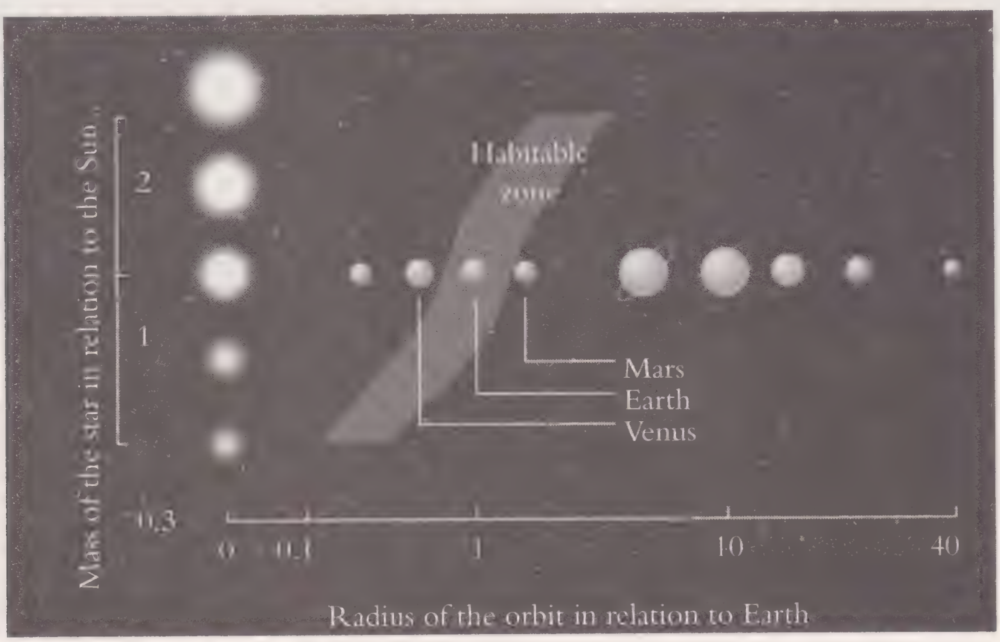
Planet Fomalhaut b inside a cloud of interplanetary dust from star Fomalhaut in an image from the Hubble Space Telescope (source: NASA).

Jupiter's density is used to calculate the diameters of the exoplanets, or the Earth's density in the case of rocky, terrestrial exoplanets. For example, to determine the diameter of Gliese 581 d (a terrestrial exoplanet), it is supposed that $\rho = 5{,}520$ kg/m^3 (the density of the Earth). The result appears in the table on page 60.

The diameters of the first multiple planetary system discovered around a star of the principal sequence, Upsilon Andromedae, are calculated in a similar way. It is composed of three planets, all of them similar to Jupiter – planets Ups And b, c and d. Their diameters, supposing that $\rho = 1{,}330$ kg/m^3 (the density of Jupiter), also

appear in the aforementioned table. By using the previous results and the periods of rotation of the exoplanets around their central star, calculation can be made of the central star's mass by using Kepler's third law; the constant a^3/P^2 is equal to the central mass of the star (see Appendix).

Many of the exoplanets in orbit revolve around their stars at a much shorter distance than any planet in our Solar System orbits the Sun; in other words, many planets are closer to their star than Mercury is to the Sun. This means that their temperature is very high. The inner part of the Solar System is occupied by the small rocky planets and the first of the gaseous and giant planets, Jupiter, is 5.2 AU from the Sun, while the extrasolar systems provide many examples of large planets revolving much nearer to their mother stars. These differences are thought to be due above all to a sampling bias, a consequence of the observation methods used. Thus, the radial velocity method used to detect exoplanets is more sensitive when the planets trace smaller orbits and are more massive. But it is clear that we can suppose that most exoplanets move in much larger orbits. It seems reasonable to expect that in the majority of extraplanetary systems, there are one or two giant planets with orbits comparable to Jupiter and Saturn's orbits in our Solar System.



Planetary habitability zone, where life may exist in exoplanetary systems and in our own Solar System.

Let's look now at the habitability of the exoplanets. Rough calculations indicate that the habitable zone of the Solar System, where liquid water can exist (temperature range from 0° to 100° C), stretches from 0.56 to 1.04 AU. The internal boundary

of this zone is between the orbits of Mercury and Venus, and the external border is just outside the Earth's orbit. Only two planets in the Solar System, Venus and the Earth, are in the inner habitable zone (the grey zone in the illustration on the next page) and, as we know, only the Earth is inhabited, as the temperature on Venus is too high, due solely to the severe greenhouse effect it is subject to. At present, Gliese 581 d appears to be the best example that has been found of a terrestrial exoplanet that orbits close to the habitable zone around its star and is a potential candidate for extraterrestrial habitability. Gliese 581 c is also in a position that could place it in the habitable zone of the host star. It has been determined that this exoplanet could have liquid water and offer the chance of life, though some studies indicate that it may be subject to a runaway greenhouse effect similar to that of Venus.

There are still many unanswered questions about the properties and characteristics of the exoplanets, with a great deal still to be studied and researched.

Chapter 3

Eclipses and Transits: Meeting Points

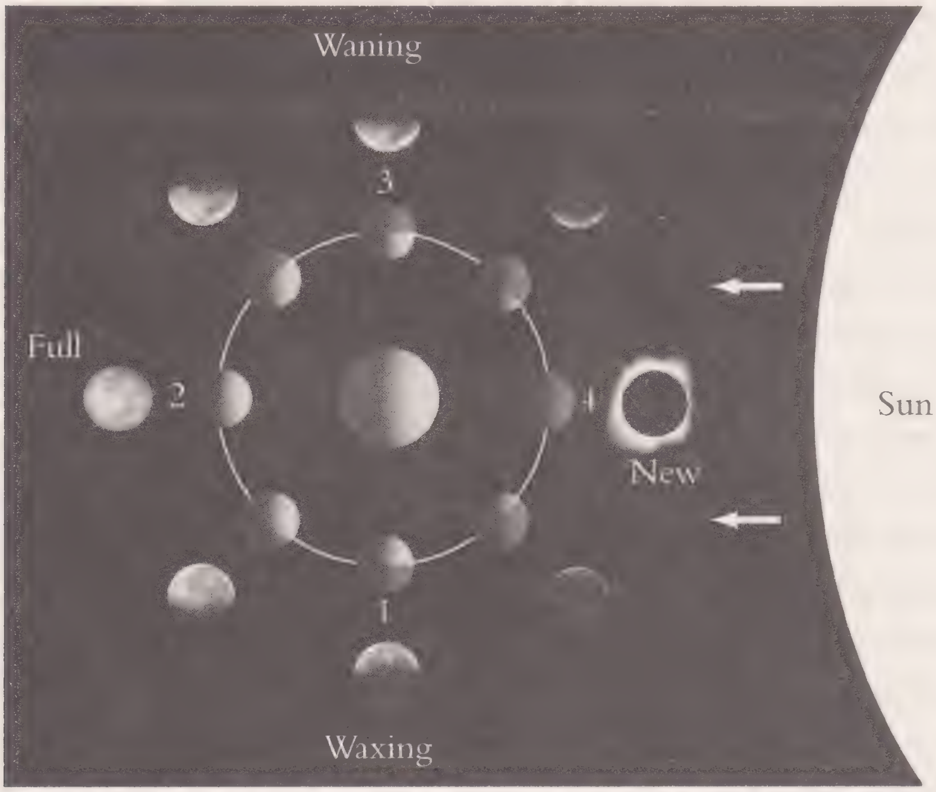
In ancient times, total eclipses of the Sun were thought to be extremely dangerous events. It was believed that the destiny of the world depended on the eternal, divine stars, and the destruction of the most important one would bring the world to a sudden end. That is why many cultures had religious rites aimed at protecting the Sun god. It became essential to find the rules that would enable solar events to be predicted, and systematic observation became necessary, which was the first step towards a scientific approach.

But eclipses have also been used to determine the relationship between distances. As we said above, Aristarchus made use of a lunar eclipse to determine distances in the Earth-Moon-Sun system. In particular, the transits of Venus were used in the past to calculate all the distances of the planets in the Solar System. Eclipses did, in fact, help humankind to reach new heights of scientific knowledge, and some of these achievements are summed up in this chapter. The study of eclipses is nothing more than the solving of a mathematical problem, a problem of positional astronomy.

The geometry of eclipses

The term 'eclipse' covers different types of phenomena, and is applied in particular to eclipses of the Moon and the Sun. In general terms, it is not easy to understand the different layout of the bodies in these two cases and it is not unusual to find newspaper reports on eclipses in which a few gaffes crop up. For example, it may not take too much effort to understand that solar eclipses always take place at a new moon, while lunar eclipses happen during a full moon. At first it is not easy for those who are not versed in the subject to see the link between lunar phases and eclipses of the Sun or to understand why there are many more eclipses of the Moon than of the Sun. We shall explain those issues in a way that is simple to understand.

In the next illustration, the phases of the Moon are shown in abbreviated form. A lunar eclipse takes place when the Moon enters the cone of the Earth's shadow and this happens whenever the satellite is in the position shown in the lower diagram. As the shadow cone is larger than the Moon, it is relatively easy for an eclipse of this type to occur. It is obvious that a lunar eclipse can only take place when there is a full moon, as the Earth has to be situated between the Sun and the Moon.



When the Moon revolves around the Earth, half of the satellite always remains lit up (small Moons in the inner circle), but from the Earth the Moon is not always seen in the same way, as can be seen by the Moons that form the outer ring. In position 1 we only see the Moon in its first quarter; in 2 the full moon is seen from Earth; in 3 it is only seen in its last quarter; and in position 4 the new moon is not seen during the day at all as the Sun's light dazzles us.

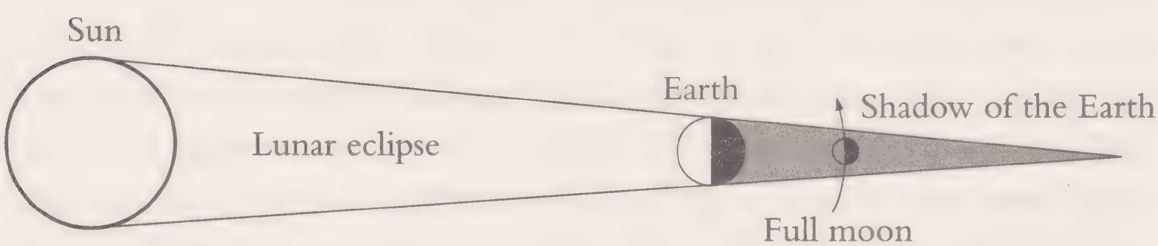


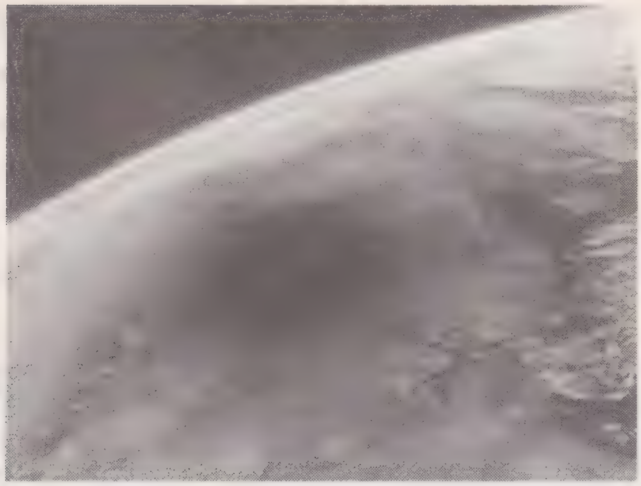
Diagram of an eclipse of the Moon (it happens only during a full moon).

Photographic montage of an eclipse of the Moon. The Moon is crossing the shadow cone cast by the Earth.

An eclipse of the Sun happens in the opposite case to the one given above. It occurs when the Moon positions itself between the Sun and the Earth, that is, when the Moon is at position 4 in the illustration on the previous page. The eclipses of the Sun can only be observed from a specific place on the terrestrial surface, which does not happen with lunar eclipses. Furthermore, for any terrestrial observer, solar eclipses occur less frequently than eclipses of the Moon.

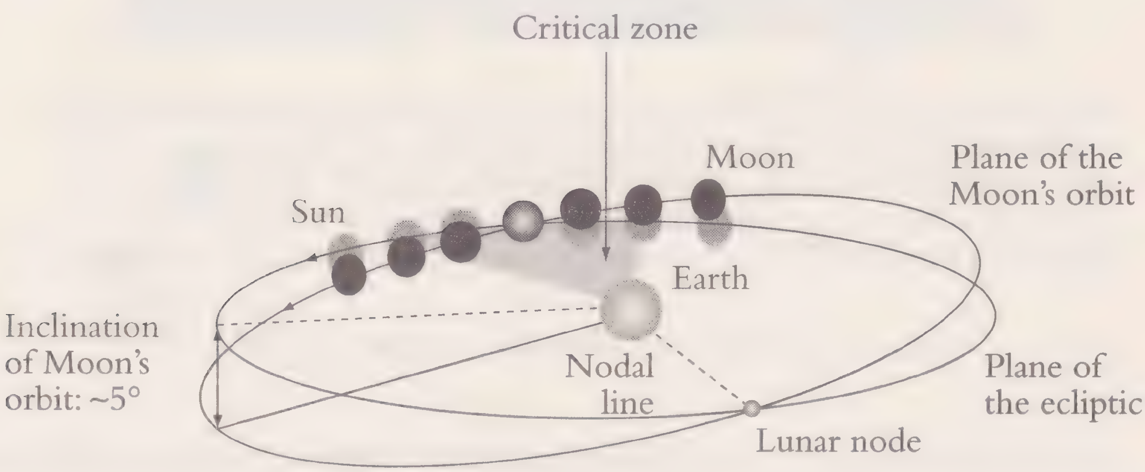
Diagram of an eclipse of the Sun (it always happens at new moon).

As can be clearly seen in the diagram, the Earth is aligned with the Sun and the Moon, both bodies being on the same side. So, the corresponding phase is new moon. It is also clear that eclipses of the Moon can be seen by far more people than solar eclipses, which can only be seen very locally. We can enjoy an eclipse of the Moon practically every year, but we may have to wait many years before getting the chance to observe an eclipse of the Sun without having to travel to see it.



A photograph taken from the International Space Station of the eclipse of the Sun in 1999. The shadow of the Moon can be seen on the terrestrial surface.

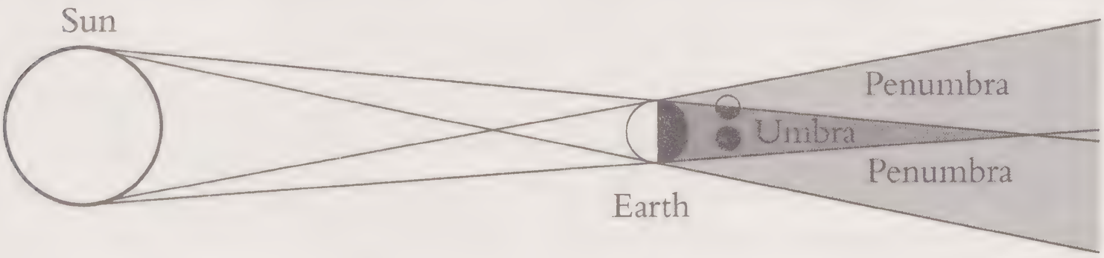
In view of the above, it might be thought that every time there is a new moon there should be a lunar eclipse, and with every new moon, a solar eclipse takes place. In that way we would have an eclipse of the Moon, and one of the Sun (albeit only partial) at every lunation – approximately every month. However, we know full well that this isn't the case. It is due to the fact that the Moon's orbit is inclined 5° with respect to the ecliptic plane, so the eclipses only take place when the Moon is close to the nodal line – the line of intersection between the Earth's orbital line round the Sun and the Moon's orbital plane round the Earth.



The nodal line is the intersection line between the Moon's orbital plane around the Earth and the plane of the ecliptic (the plane of the Earth's orbit around the Sun). There can only be eclipses when the position of the Moon is close to the nodal line, that is to say, when it close enough to the plane of the ecliptic for a shadow to be cast upon it, or by it.

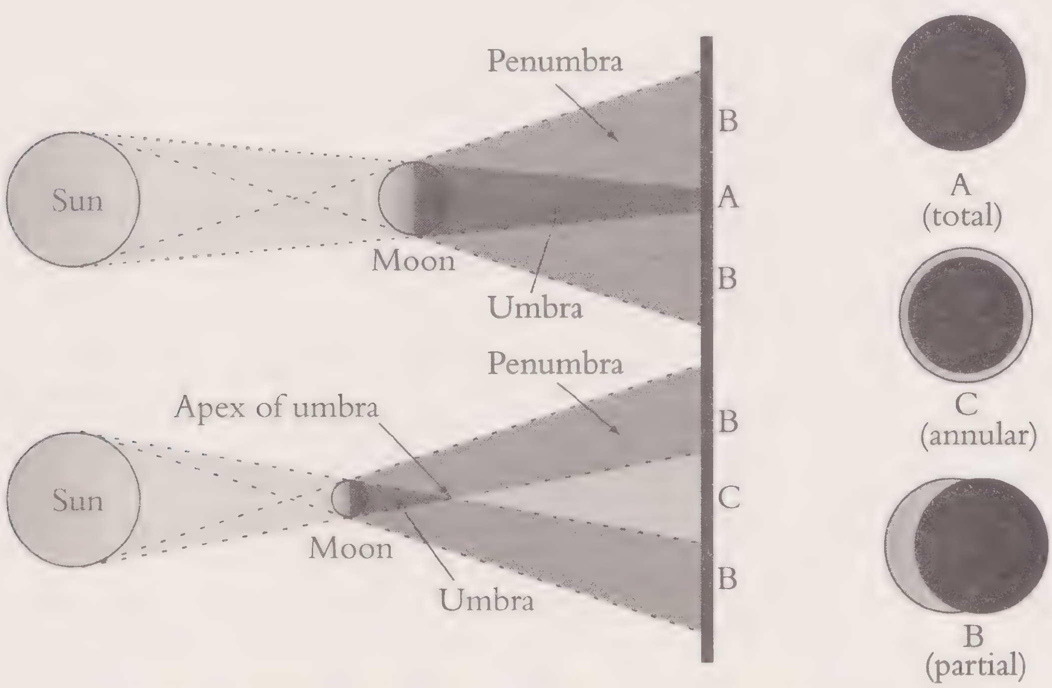
The umbra and penumbra zones

In any eclipse, two zones can be established geometrically where there are shadows, the umbra, with a dense shadow, and the penumbra, with a less dense shadow.



The external tangents bound the umbra, and both these and the internal tangents bound the penumbra. Total lunar eclipses take place when the Moon is totally inside the umbra zone; if only a sector of it is inside, then it is a partial eclipse.

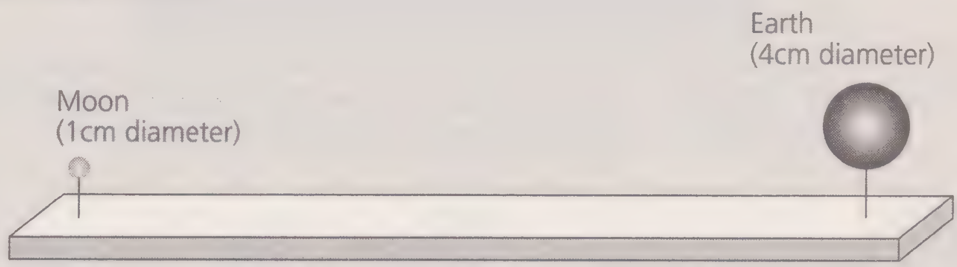
What's more, in the case of eclipses of the Sun, we have to differentiate between a couple of other situations depending on the relative distances of the Earth and the Moon. The orbits of the Earth around the Sun and the Moon around the Earth describe slightly eccentric ellipses. For that reason the distances are not always the same, as they would be if it were a circular orbit, and on occasions the Moon does not obscure all the solar surface. So, there are total eclipses and annular ones, ie when the Moon does not manage to cover all the Sun's diameter and a 'ring' of light is left in view.



Depending on the Moon's distance, eclipses may be total or annular.

THE EARTH–MOON MODEL

To help in understanding the phases of the Moon and the different types of eclipse, you can make a simple model, like the one shown in the illustration below. To construct it, hammer two nails 3cm or 4cm long into a strip of wood 120cm apart (the piece of wood needs to be about 125cm in length). On each nail, fix a sphere made out of polystyrene, one 4cm in diameter and the other 1cm, which will simulate the Earth and the Moon respectively. It's very important to follow the measurements given so that the model is to scale (1:320,000), thus maintaining the proportions of distance and diameter.



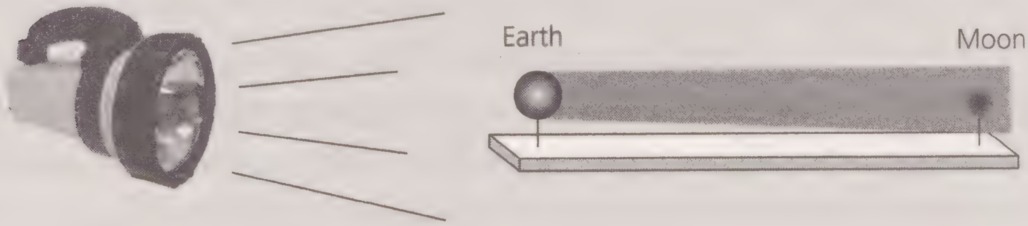
Distance between nails: 120cm

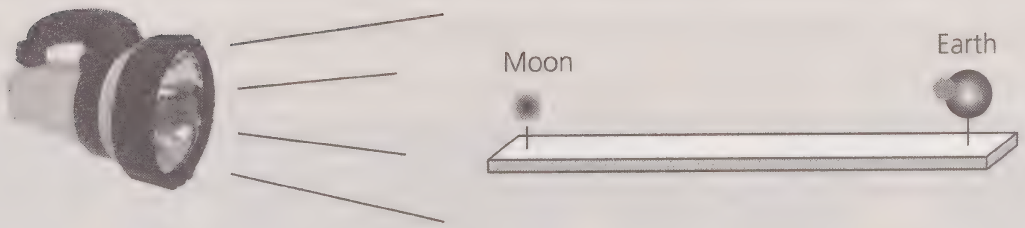
To simulate the phases of the Moon, take the model to a sunny place where the Moon is also visible. The model is used by situating the larger sphere (which represents the Earth) close to us, and the small one (representing the Moon) in the direction in which the Moon is situated in the sky. We can see both moons (the real one and the simulated one on the model) at the same size and in the same phases. If the orientation of the base is changed, the different phases of the Moon will be reproduced depending on the light that the small sphere receives from the Sun. The Moon has to be moved from right to left to reproduce the phases in order.

It is better to do all this in the open air, but if it is cloudy it can be done in a room by using a torch to take the place of the Sun. The light from the torch must be intense and focused into a beam.

To simulate the eclipses, we could use the model in the open air provided that it is a sunny day, otherwise inside a room with a light taking the place of the Sun. We take the model and situate the sphere of the Earth facing towards the Sun, and we move the Moon within the shadow cone produced by the Earth (top diagram, opposite page). As the cone is larger than the Moon, an eclipse can be reproduced (with the shadow of the cone covering the surface of the Moon). Furthermore, it is clear that the Earth is aligned between the Sun and the Moon, so the corresponding lunar phase is a full moon. By using the Sun or the torch again, a solar eclipse is obtained by situating the model

moon against the light. The model has to be moved until the shadow of the Moon appears on the surface of the sphere representing the Earth (lower diagram, below).





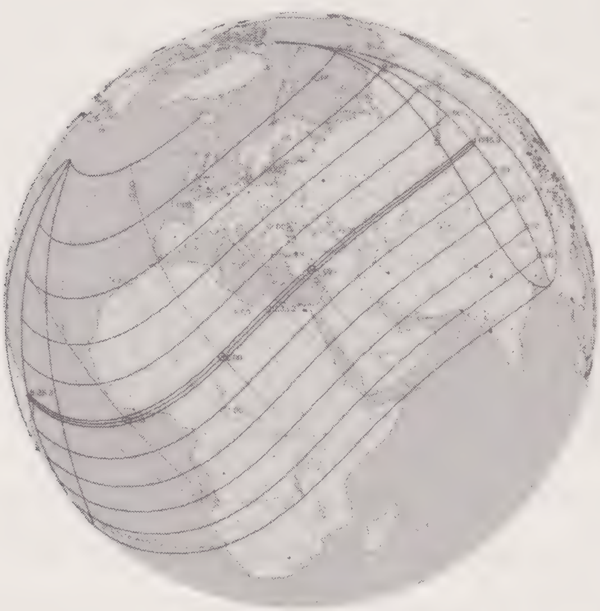
Simulation of an eclipse of the Moon (top) and of the Sun (represented by the lamp, above).

Anyone who tries this experiment will see that it is not so simple to get this type of eclipse as the previous one. What is more, it is clear that the eclipses of the Sun can only be seen from the part of the terrestrial surface where the Moon's shadow is located. That is not the case with lunar eclipses. Therefore, it is now obvious that the Sun's eclipses take place far less frequently than those of the Moon. As can be clearly seen in the illustrations, the Earth is aligned with the Sun and the Moon, with both bodies on the same side; so, the corresponding phase is that of the new moon.
After seeing this, it is easier to understand why it is normal for us to be able to observe an eclipse of the Moon practically every year, but that we have to wait many years to get the chance to witness an eclipse of the Sun near to where we live. If we slowly move the shadow of the Moon over the surface of the Earth, we can visualise the 'shadow line' that it leaves on the surface of the planet (see page 72).

A great show: a total eclipse of the Sun

I imagine that the reader will have had the chance to witness an eclipse of the Moon, and may even have attended a partial or annular eclipse of the Sun. I personally believe that a solar eclipse is more spectacular than a lunar eclipse; in fact, I don't think there is anything to match a total eclipse of the Sun. The first one I had the chance to see took place on 11 August 1999. I was in the north of France, near the border with Germany, at Briey. The clouds had us on edge all morning, but in the end we were able to see the eclipse, though it was a little hazy due to the cloudiness. I remember perfectly that in the time leading up to the total eclipse, when there was a marked decrease in luminosity, a great black bird began to fly in an enormous spiral, soaring up dizzily high till it faded into a little black dot. I will never forget it. The poor bird was in total panic mode, and no wonder. The eclipse did not last more than three minutes, but three minutes of night-time at mid-morning is a long time when there is no certainty as to whether the Sun is ever going to reappear or not.

In March 2006, I had another opportunity to witness another total eclipse of the Sun. It was in Sallum, Egypt, in a part of the desert near to the border with Libya. The Egyptian government organised the infrastructure to receive thousands of journalists,



The total eclipse of the Sun on 29 March 2006, began on the coast of Brazil, crossed the Atlantic, touched on the African coast in Ghana and Togo, crossed Africa reaching its maximum duration of totality in the Sahara (4 minutes and 7 seconds), passed over the border of Libya with Egypt at the city of Sallum, crossed the Mediterranean towards Turkey, crossed the Black Sea entering into the Russian Federation, and finished its trajectory in Mongolia.

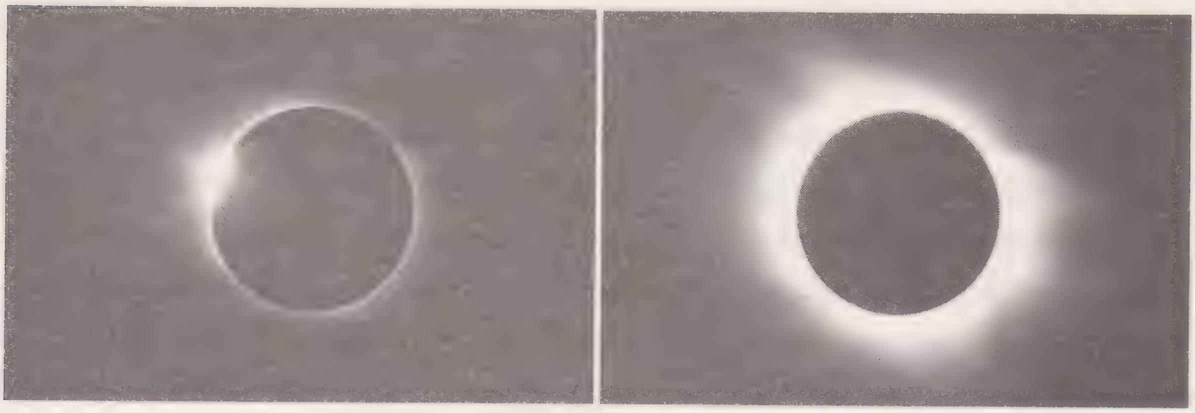
scientists and visitors interested in witnessing the event. The area chosen was in the middle of the desert. It only had four buildings: a restaurant, a couple of army buildings and a mosque. In true Bedouin style, a number of temporary shops were set up to cater for the visitors. Some folk had arrived the previous day, others, like me, had arrived that same morning. The desert appeared to be covered in mist and the Sun could hardly be made out. As the Sun rose up higher over the horizon, the mist began to disappear and, in the end, we were able to enjoy fine, bright weather.

A total eclipse is preceded and heralded by a certain process. About an hour before, the partial eclipse begins. With the protection of filters, we gazed up at the Moon's rugged limb while the light grew dimmer and dimmer as the total eclipse drew near; the animals, and we people too, felt that something was going to happen. Everything around us grew darker and darker, and the temperature fell. The level of light drops even more and it becomes difficult to distinguish the things and people around you. Complete darkness arrives suddenly – the total eclipse has begun.

The change of light levels in a solar eclipse is impressive. These photographs, taken from the same spot, show the progression. The top row of photographs shows phases of an eclipse. The bottom row of photographs shows the level of daylight on the ground during each phase.

The observation in Sallum was superb. Minutes before the total eclipse, we saw a splendid 'diamond ring' and Baily pearls (solar light that filters through the irregular border of the Moon), and finally, the solar corona with sharp, whitish crests that were incredibly beautiful. Through telescopes fitted with filters it was also possible to distinguish some solar prominences that were visible on the Sun's surface. After the

four minutes of total eclipse, the process was reversed; again the spectacular 'diamond ring', the increase of luminosity, and again we enjoyed a partial eclipse that, little by little, began to let the light come through again.



Above left: the effect known as the diamond ring. Above right: the corona.

A total solar eclipse is one of those things that you never forget as long as you live, but it has to be seen live. Any image – a photograph or a movie – that we might see after the event is simply a 'decaffeinated' eclipse, in the sense that it lacks punch. It is like comparing the actual sight of a fire crackling away in a fireplace with a photograph of one: there's something missing somehow. The real fire is perceived in its entirety, not only through the eyes: our whole body feels it, the skin, the ears... while a photo gives us a limited, restricted and altogether poorer sensation. After living through such an experience, one can understand the terror that this type of phenomenon aroused in ancient peoples. In our case, we have the certainty that the Sun is going to be 'reborn', but for a people with no written records, witnessing one of these phenomena, that no one alive had seen before, terror and panic must have abounded. For those who love discovering something new, I would like to recommend they do not miss the next eclipse of the Sun. It's well worth making any journey for!

Eclipses and scientific progress

Throughout the history of humankind, predicting lunar eclipses did not prove to be so difficult, but for total eclipses of the Sun, events that happen very rarely at a local level, it was different. As said above, the Moon's orbital path is slightly inclined

– about 5% – with respect to the Earth's orbital plane; without that inclination we would have a lunar eclipse every month, coinciding with every full moon. But in most cases the Moon passes by a little before or a little afterwards, and there is no eclipse. However, when it is near enough to the line of intersection it does happen, which is normally twice a year.

Taking into account the chances of its being night-time in a particular place when that happens, there is an average of one lunar eclipse per year. It was clear that the phenomenon repeated every 19 years. That was the origin of the Saros Cycle, which, surprisingly, was already known back in the Stone Age. The ruins of Stonehenge, in Wiltshire, feature a ring of 56 holes which have been dated, at their oldest part, to 1900 BC (with modern data $3 \cdot 18.61 = 55.83$ years, which is only out by one day per year, which is truly extraordinary for a calendar of that era).

Solar eclipses are much more difficult to predict. Thus it came about that two Chinese astronomers were sentenced to death for failing to predict a solar eclipse, though the Greek philosopher, Thales of Miletus, was successful in predicting a total eclipse visible from Greece. It took place at the same time as a battle between the Persians and the Medes in the year 585 BC in what is now Turkey. This 'divine intervention' brought about an armistice, and Thales became a renowned figure.

The eclipses provided decisive backing for the first heliocentric model proposed by Aristarchus. From observations, Aristarchus deduced that the Sun's diameter was 19 times greater than the Earth's, and it seemed pointless to argue that a large body would revolve around a small one, which led him, intuitively, to establish a heliocentric model which was basically correct, though not totally free of errors; even Copernicus referred to Aristarchus's ideas when he proposed his system. Again it can be seen that it all comes from an intelligent analysis of the observation of eclipses.

Another problem that was solved by using eclipses was the determination of a point's longitude. Geographical position as far as latitude was concerned could be calculated in several ways. In the northern hemisphere one simple method was to measure the height of the North Star. This method was known to mariners, as was the diameter of the Earth, which enabled them to easily measure north-south distances. But east-west positions, ie longitude, and along with them, their corresponding distances, created serious problems for centuries. It was, in fact, a lunar eclipse observed by Alexander the Great in India that helped to solve this problem, and by

coincidence. When his expedition arrived back in Greece, the people explained that the eclipse had occurred a few hours before midnight. As sunset in Greece is a few hours before sunset in India, it became clear how many degrees to the east India is situated from Greece. In this way, they were able to calculate the distance to India in units of longitude.

The problem of the determination of longitude and east-west distances had been a challenge for centuries. America had been discovered but how should, for example, the distance to Mexico be measured? It would have been easy to work it out by using synchronised clocks, but clock technology was still not well-developed. The pendulum clock had been invented by Huygens, but how could such a clock be taken on a ship? How could a clock rocked around on a sea voyage remain synchronous with one standing on land? It seemed impossible.

Again, it was an eclipse that helped solve the problem. In this case it was the eclipse of Jupiter's moons discovered by Galileo. It was he himself who proposed using that eclipse as a pointer that could be simultaneously read in many countries. The moment that a Jovian moon suddenly disappears on entering the shadow of Jupiter could be 'read' simultaneously in Europe and Mexico, and this method would enable the hour to be accurately determined to within one minute. But, as we shall see later, the solution had to wait until sufficiently accurate clocks could be built.

It was to be Giovanni Cassini at the Paris Observatory who initiated the project to be able to read such global time. But when they came to evaluate the data taken from the observation, some difficulties emerged. The time between the two eclipses varied by 15 minutes. Roemer, a young Dane in his team, explained this lag as being caused by the time needed by the light to reach the Earth in the different positions on its orbit, which is not always at the same distance from Jupiter. And with it, at the same time, he devised a method for measuring the speed of light.

According to the Newtonian theory of gravitational attraction, the force of attraction of the Sun could bend the rays of light from the distant stars and this deviation should be of 0.875 seconds of arc. But according to Einstein's theory of relativity, this deviation was double. It was precisely during the solar eclipse of May 1919 that Sir Arthur Eddington measured this effect and confirmed Einstein's prediction. The British astrophysicist, after carefully studying data from observations, obtained a value of 1.98 seconds of arc. Another type of 'eclipse' enabled Einstein's theory to be proven with more accuracy. It was the occultation, in 1987, of a quasar by the Sun, which allowed measurement of the

deviation of the quasar by interferometry, providing verification of the theory with an error of 0.1%.

By taking the term eclipse in its most general sense of 'concealment', nowadays eclipses can again be said to be at the forefront of scientific research. One way of detecting material in the Universe, in particular brown dwarfs, which are too dim to be detected by means of their own light, is a consequence of the 'microlensing' effect that occurs when a brown dwarf is hidden by another star. The light from the dwarf star is deflected and focused along the optic axis of the gravitational lens and for a short space of time gives off an enhanced gleam that allows it to be detected. In this same field, the enormous masses formed by a galaxy and cluster of galaxies deviate the light from other objects situated a great distance beyond. Since 1979 when the first gravitational lensing was discovered, this has been a very important issue in modern-day astronomy, and one that we shall deal with later.

Einstein had predicted this effect in 1912 in some of his notes, but he did not publish as he considered it to be of little importance. However, after being urged on by a friend, he published in 1936 with the intention, he said, of "making this poor boy happy". Today it is one of the primary fields of research in astronomy.

Gravitational lenses: they don't eclipse, they magnify

Einstein predicted gravitational lenses – the fact that a star situated in the foreground could magnify the image of a star situated in the starry background. But he was sceptical as to whether this illusion could ever be seen, and thought it to be a hypothesis that was too improbable to be of practical interest. Today, astronomers are using this effect to be able to see much further. It is, in fact, as if the Universe itself were equipped with it own superb telescopes enabling us to see what has never been seen before in space and time. The study of these lenses can still be considered as a very new observational science.

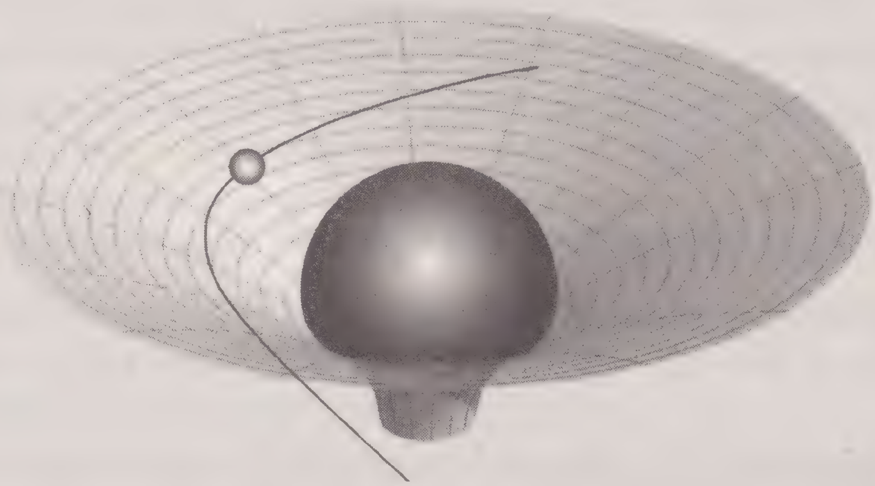
Light always follows the shortest possible path between two points, but if a mass is present, space curves, and in this case the shortest path is a curve. This idea is not so difficult as it might appear at first, as we can find parallels with the surface of the globe, upon which the shortest distance between two points always describes an arc.

In general terms, we can imagine gravitational lenses as ordinary lenses, but in the case of the former the deviation of light is caused by mass instead of refraction. The most important difference is that an ordinary convex lens has a

single focal point while a gravitational lens does not; instead, it focuses light on a zone, not on a point.

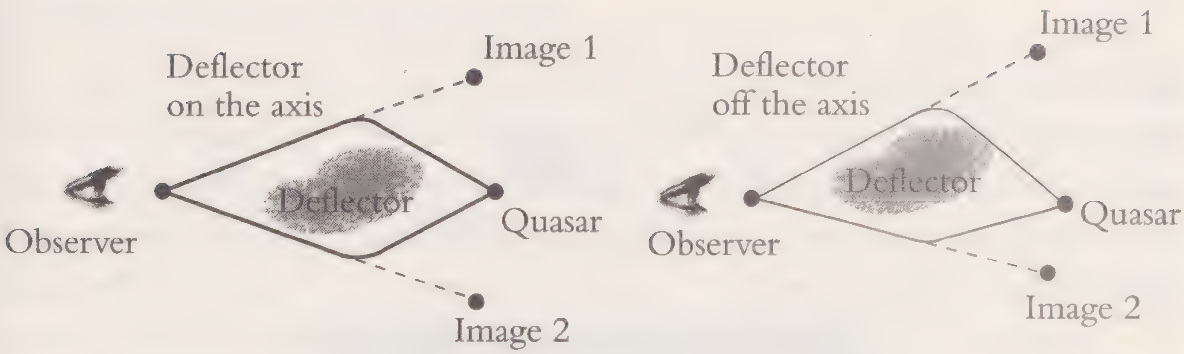
DEFLECTION OF A SUN RAY DUE TO THE CURVATURE OF SPACE

It is very simple to simulate the curvature of space caused by a black hole by using a piece of slightly elasticated cloth and placing a heavy ball in the centre. If we roll in a lighter ball, its trajectory will follow a curve in the space. This is simulating the trajectory of a light ray, which follows a curve and not a straight line, as shown in the illustration. The degree of deviation depends on how near the light ray passes to the central body and how massive this central body is. The angle of deviation is directly proportional to the mass and inversely proportional to the distance.



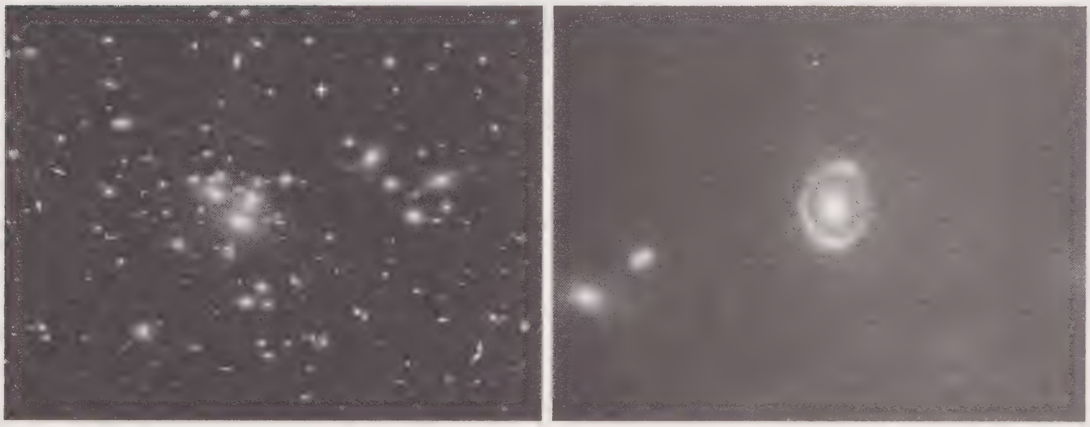
Basically, gravitational lensing produces a curve in light rays. The result is that objects seem to be in a different place and appear magnified. As they have no focal point they are not perfect lenses and the images they produce are distorted. They may generate distorted images or multiple images of an object.

The deviation may cause an apparent change of position of a star, galaxy or quasar in the sky. It can also change the size and magnify objects. Some observers have noted magnifications of more than 100 times.



Gravitational lenses change the apparent site of a star, galaxy, or quasar. A good example is the famous eclipse of 1919, which proved Einstein right.

If the light subject to deviation is from a galaxy, cluster or other astronomical object without a single-point structure, the images obtained are a set of bright arcs, such as the spectacular arc of the Abell Cluster (photograph, below left). Sometimes, if the lensing system is perfectly symmetrical, the rays converge and the resulting image is a ring, called an Einstein ring (photograph, below right); the bright point in the centre of the bull's eye is the nearest galaxy.



As gravitational lenses have no single focal point, they are not perfect lenses and can produce multiple images, as shown in the image on the next page. The examples of multiple images of quasars in the shape of what is known as the Einstein cross are the best known.

As gravitational lenses have no single focal point they can produce multiple images. The photograph above shows a quasar known as an Einstein cross (source: NASA).

GRAVITATIONAL LENSING WITH THE BASE OF A WINE GLASS

We can simulate gravitational lensing by looking through the stem of a wine glass, something that is easy to get hold of, but that needs the stem cutting off by a glazier! If we place it on a sheet of graph paper and look through it we can see the deformation shown in the photograph.

By moving the stem slowly from right to left over an object, for example, a dark circle that simulates an astronomical object, we can recreate the different real objects observed: segments of an arc, the Einstein cross and ring, as can be seen in the photographs on the next page.

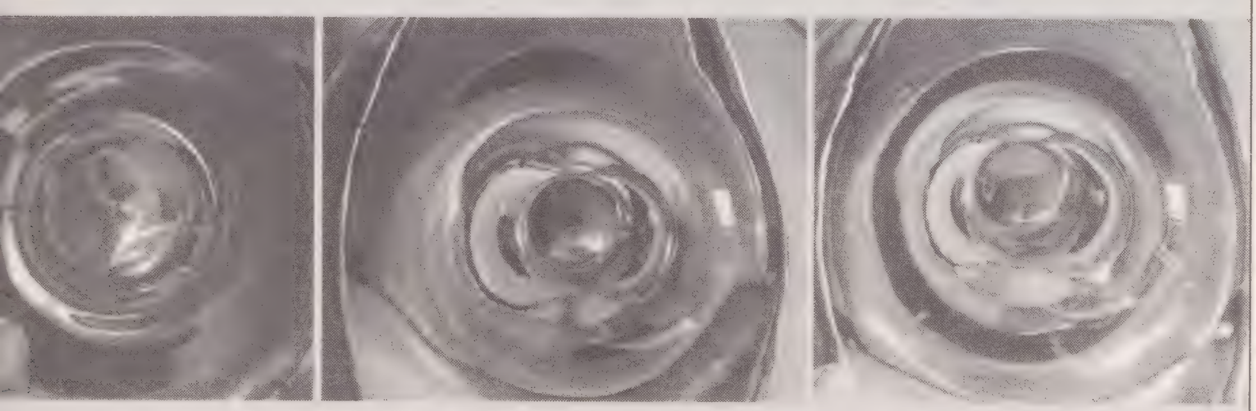
We can also simulate the deformation with a clear wine glass, by looking from above, while for the Einstein ring or multiple images, a glass made of red glass can be used with an LED light situated on the other side of the glass so that

Seeing further in time and space

As well as the parsec and the astronomical unit, light years are also used as units of distance. Nowadays, this unit of measurement is not often used by professionals, but it is nevertheless very intuitive, and useful for putting across and demonstrating something that is quite surprising. When we look at the starry sky we are not seeing the present-day image but rather a set of objects that look like they did in the past. And what's more, there is a mixture of objects from different eras in time.

As is well known, the speed of light is $c = 300{,}000$ km/s; therefore, the distance that corresponds to 1 light-second is 300,000 km. For example, the time that light takes to travel from the Moon to the Earth, which are 384,000 kilometres apart, is $384{,}000/300{,}000 = 1.28$ seconds. Using these units, the time required for light from the Sun to reach Earth is 8.3 minutes.

From the southern hemisphere (it is not visible from the northern hemisphere) it is possible to observe Proxima Centauri, the star nearest to Earth, which is 4.3 light years away. Sirius, the brightest star seen from a large part of the northern hemisphere, is 8.6 light years away. In either of the two cases it is obvious that the light we see has taken several years to reach our eyes but, for other objects, that time can be far greater.



Arc segment *Einstein cross* *Einstein ring*

the light ray passes through it. By moving it from right to left and up and down we can see how the light produces repeated images, and in some cases, arcs. This is because the glass acts as a lens which 'deforms' space. In particular, we can at times observe an amorphous shape, four red points or an arc between two red points.

The Orion Nebula, known as 'the most beautiful', is a stellar nursery where some 700 stars are now being created. This cloud is situated 1,500 light years away, so that means the image of it that we can now see is as it was around the time of the fall of the Roman Empire (476 AD, when Romulus Augustus was deposed).

The spiral galaxy of Andromeda can be seen with the naked eye – it is one of the closest galaxies in our neighbourhood and looks like the Milky Way, so we can imagine we were seeing our own galaxy from far beyond its edge. Andromeda is about 700 kpc away, in other words, more than 2 million light years, so we see this galaxy as it was when the first hominids were walking about on the Earth.

The Orion Nebula in an image taken by the Hubble Space Telescope (left) and the Andromeda galaxy (source: NASA).

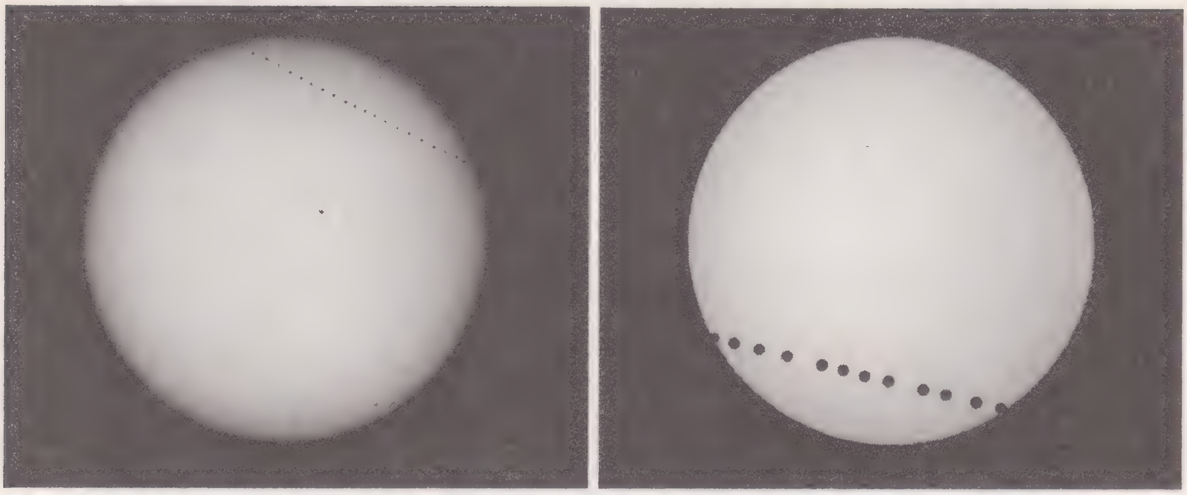
It may turn out that when we are observing an object today, in reality its properties have changed completely. For example, the supernova SN 1987A, whose great explosion was observed in 1987 in the Large Magellanic Cloud, was not seen before as a really bright object from the Earth, but the explosion had already taken place. It being 51.4 kpc away, the view of the explosion had taken 168,000 years to reach us. But if we take into account the way that gravitational lenses see very, very distant objects, which are many more light years away – several billions in fact – we can peer into the depths of time.

Supernova SN 1987A in the Large Magellanic Cloud during the explosion, and four years later, when the echo can still be seen expanding (source: the Anglo-Australian Observatory).

Another type of eclipse: the transits of Mercury and Venus

The passing of an inner planet (Mercury or Venus) in front of the Sun is called a transit. Throughout history, transits have been very useful as they have often enabled advances to be made in astronomical research. A transit happens when the Sun, the inner planet, and the Earth are all aligned. It would be the same as an eclipse of the Sun but, instead of the Moon, what moves in front of the Sun is the planet and, as it is much further away than our satellite, it can only be seen as a small spot over the surface of the Sun. This spot moves along as time goes by, which allows repeated exposures to be taken and, once they have been superimposed, they give a very clear idea of the trajectory, as can be seen in the photographs on page 84:

As the orbits of Mercury and Venus are slightly inclined in respect to the ecliptic, a transit only takes place when those planets are near to the nodal line (the line of intersection of their orbits with the ecliptic). There are certain rules for calculating the periodicity of these phenomena, though they are quite complex. On average, Mercury transits the solar disc around 13 times a century in accordance with very complex rules.



Above left: the transit of Mercury; the planet's trajectory can be seen across the solar disc. Above right: the transit of Venus. As well as its trajectory, note the size of the planet in comparison to the diameter of the Sun (source: Agrupación Astronómica Deneb).

<table>
<tr><th colspan="3">Transits of Mercury and Venus in the 21st century</th></tr>
<tr><td rowspan="14">Mercury</td><td></td><td>7 May 2003</td></tr>
<tr><td></td><td>8 November 2006</td></tr>
<tr><td>9 May 2016</td><td></td></tr>
<tr><td>11 November 2019</td><td></td></tr>
<tr><td></td><td>13 November 2032</td></tr>
<tr><td></td><td>7 November 2039</td></tr>
<tr><td>7 May 2049</td><td></td></tr>
<tr><td>9 November 2052</td><td></td></tr>
<tr><td></td><td>10 May 2062</td></tr>
<tr><td></td><td>11 November 2065</td></tr>
<tr><td>14 November 2078</td><td></td></tr>
<tr><td>7 November 2085</td><td></td></tr>
<tr><td></td><td>8 May 2095</td></tr>
<tr><td></td><td>10 November 2098</td></tr>
<tr><td rowspan="2">Venus</td><td>8 June 2004</td><td></td></tr>
<tr><td>6 June 2012</td><td></td></tr>
</table>

The transits of Venus are less frequent: four transits occur in a period of 243 years, at intervals of 105.5; 8; 121.5 and 8 years; intervals create 'pairs' of transits. The 243-year cycle is relatively stable, but the intervals vary over time due to perturbations that the other planets cause in Venus' orbit.

When was the first transit observed?

By using Tycho Brahe's precise observation data, Kepler made up what are known as the *Rudolphin Tables*, which catalogue the movement of the planets quite well. With these data, in 1629 Kepler was able to announce a transit of Mercury for 7 November 1631, and another for Venus on 6 December the same year. He anticipated that the observation of these transits could be carried out in a dark room and by opening a small hole in a window that projected the image of the Sun onto a screen.

Some observers were able to observe Mercury's transit by placing a telescope in the hole in order to enlarge the image. That was how it was observed in Paris, where Pierre Gassendi noted that Mercury appeared to have a surprising diameter – 12 seconds – much less than expected. However, the transit of Venus was not seen as it took place after sunset throughout Europe.

Some years later, the English clergyman Jeremiah Horrocks (1618–1641), who had studied mathematics and astronomy, calculated that there would be another transit of Venus on 4 December 1639. Although the transit fell on a Sunday, he was able to rearrange his ecclesiastical tasks so as to be able to observe the transit. Horrocks saw that the diameter of Venus was less than 1' (minute of arc; the apparent diameter of the Sun is approximately 30'), but he was only able to carry out three observations (at 3.15, at 3.35 and at 3.45) before the Sun set.

Jeremiah Horrocks observing the transit of Venus with a lens in a darkened room.

Around 1640, the English astronomer and mathematician William Gascoigne placed some wires across the focal point of a telescope and made them moveable, and in this way invented the micrometer. After this the telescope went from being a qualitative tool to being a quantitative precision instrument, as the micrometer enabled very small angles to be determined. Additionally, the telescope could be coupled with a graduated circle to take other angular measurements.

In the successive editions of *Principia* and *Opticks*, Newton published different values for the Earth-Sun distance, that is, the solar parallax, values that ranged from 10" to 13". At that time, the only correct datum was that the solar parallax could not exceed 15" (the real value is 8.794143").

Knowing this value was important for the accuracy of the astronomical tables, which were not only used for astronomy but also for navigation. Furthermore, knowledge of the Solar System now allowed the relative values of distances of all the planets to be known, and the only thing missing was to determine one distance in explicit form: the solar parallax.

In this state of affairs, Edmond Halley, who had observed the transit of Mercury in 1677, promoted the idea of using the transits of Venus in 1761 and 1769 to determine the parallax. The method he proposed was to observe the transits from two places distant from each other and to note the instants in which the planet entered and exited the solar disc. They would have to determine the angular distance between the trajectories of the planet observed from the two places as a fraction of the diameter of the Sun, and afterwards this diameter could be calculated in miles and, finally, the distance from the Sun to Earth could be worked out.

So the observation required just a good telescope and a good clock. The transits of Venus were more suited to this task than those of Mercury as the latter planet is nearer the Sun, and the angular separation is less. It is quite small in the case of Venus, as it is in the order of one-fiftieth of the diameter of the Sun. Later on we shall take a more detailed look at the difficult work carried out by astronomers with the aim of observing the transits of Venus.

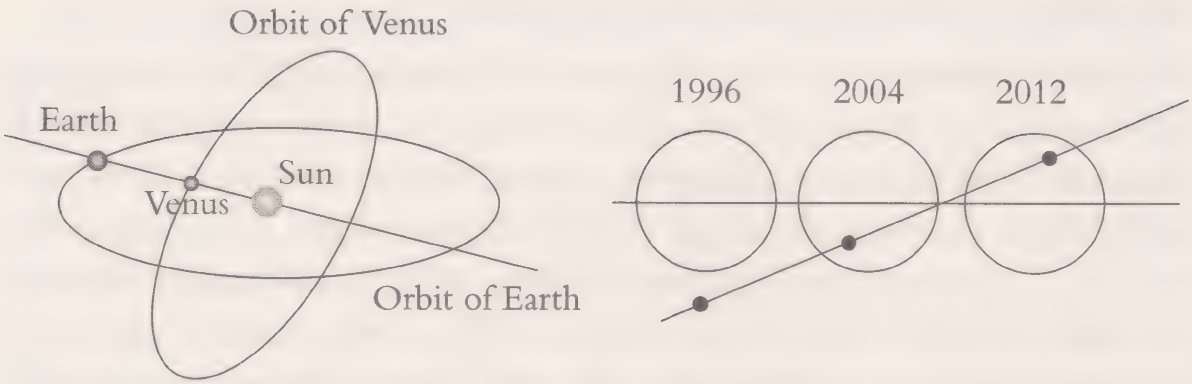
While these transits were crucial for determining the Earth-Sun distance, Mercury's transits were also of great interest. The French mathematician, Urbain Le Verrier, discovered the advance of Mercury's perihelion by studying the observations of the transits of Mercury from 1631 up to the middle of the 19th century. That revelation had many implications for Einstein's theory of relativity.

Why do the transits of Venus occur in pairs?

The period of Venus' orbit is 224.7 days, while the Earth's is 365.25 days. By dividing 365.25 by 224.7, the resulting quotient is 1.6255. So, while the Earth makes a complete revolution round the Sun, Venus makes 1.6255 revolutions; $1.6255 = 13/8$ approximately. We can therefore say that if the Earth makes n revolutions round the Sun, Venus makes $13n/8$ revolutions. When will both planets coincide? It is, obviously, when $13n/8$ is a natural number. In other words, when n is equal to a multiple of eight. So, we have deduced that every eight years, both planets must coincide on line and, therefore, it can be concluded that every eight years we shall have the chance to see a transit of Venus from some place on the globe. But, as can be seen in the following table, the reality is very different.

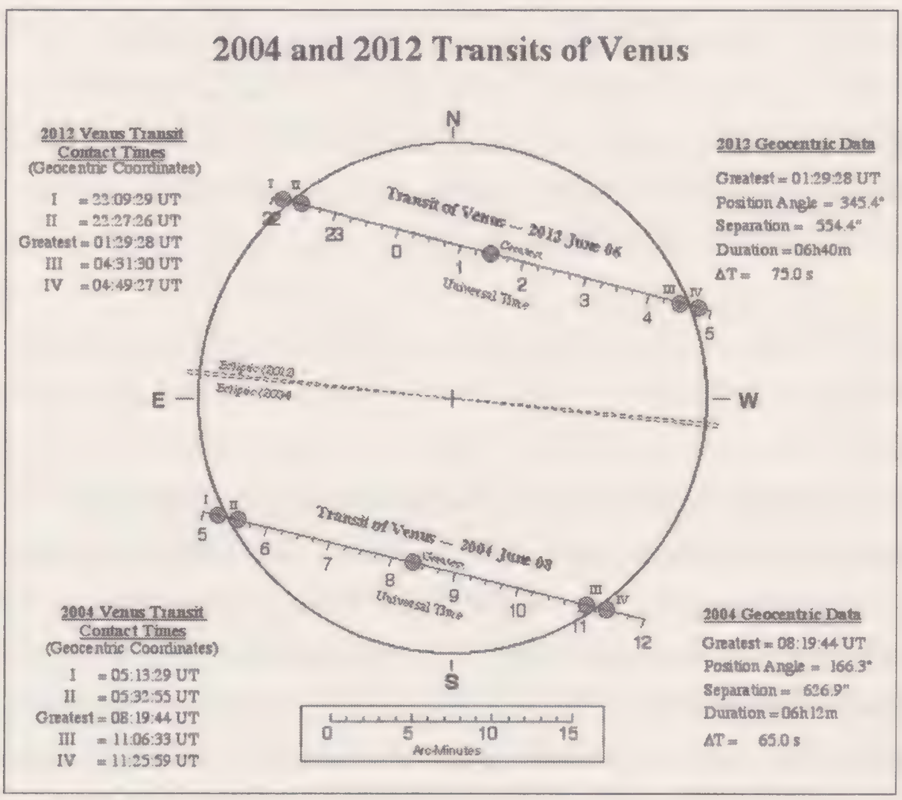
Transits of Venus observable since the 17th century	
	7 December 1631
	4 December 1639
6 June 1761	
3 June 1769	
	9 December 1874
	6 December 1882
8 June 2004	
6 June 2012	
	11 December 2117
	8 December 2125

It can be seen that when there is a transit, after eight years there is another, but the table shows us that it is necessary to wait more than 100 years for another pair of transits – still separated by intervals of eight years. The question is, then, what is the reason for this? The answer is simple. The first calculations would be true if the plane of Venus' orbit and that of the Earth's orbit (the ecliptic plane) coincided. But the truth is that the plane of Venus' orbit is inclined 3.4° in respect to the Earth's plane. Consequently, Venus will only cross in front of the Sun when both planets are near to the nodal line, ie the line of intersection of the two orbital planes, provided that the distance between both orbits is less than the solar diameter.



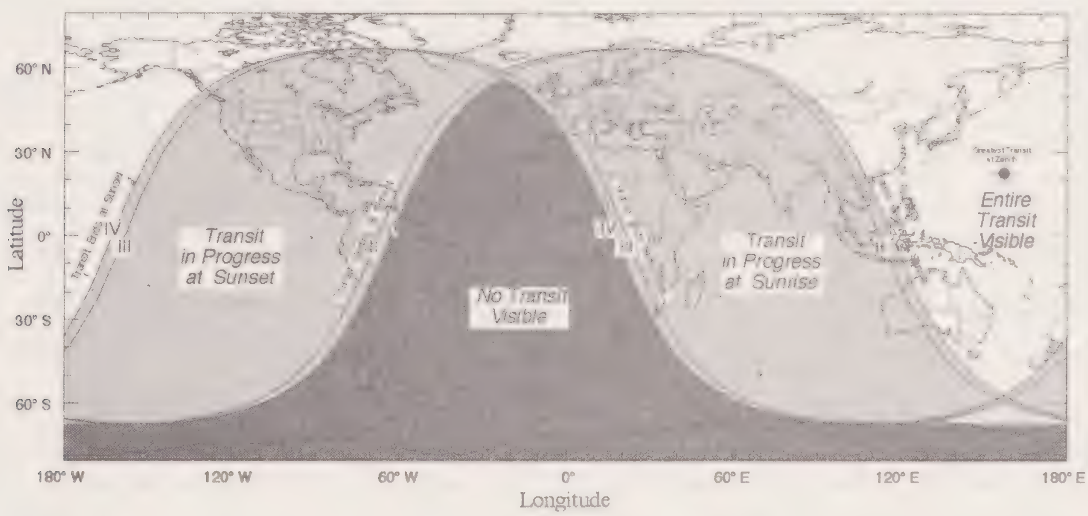
Orbits of the Earth and Venus (above left) and the trajectory of Venus across the Sun (above right).

So, for example, transits were visible in 2004 and 2012, but not in 1996 because Venus was too far out of the ecliptic plane. We will, therefore, have transits when Venus and the Earth coincide in each of the ascending and the descending nodes. Venus and the Earth meet on two occasions (separated by eight years) in the ascending node in December, and after 121.5 years they meet again twice in the descending node in July. After 105.5 years they meet up again a couple of times in the ascending node, with the cycle being repeated indefinitely.



The path of the transit of Venus on 8 June 2004 and 6 June 2012 (source: Fred Espenak, Nasa/GSFC).

It must be mentioned that the transits will not always be visible from any part of the terrestrial globe because, obviously, they have to take place when it is daytime. In other words, the Sun has to be above the horizon for us to see a transit. For example, the transit in 2004 was visible in all of Europe, but the one in 2012 excluded much of the Atlantic coast. (Although technically visible, it was so close to sunset it was very hard to observe.)



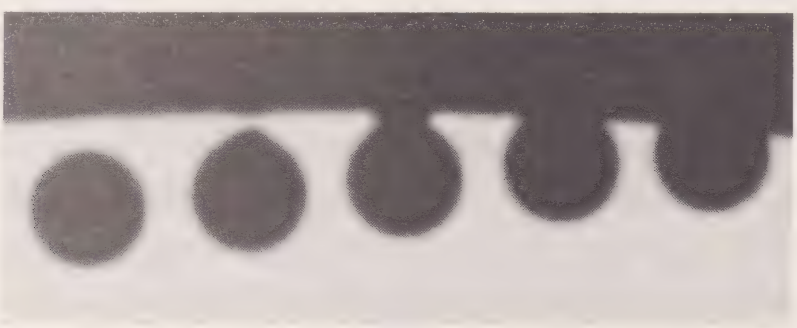
World viewpoints for the transit of Venus of 6 June 2012 (source: Fred Espenak, NASA/GSFC).

The campaigns of the 18th and 19th centuries

Although Horrocks considered the possibility of using the transit to calculate the distance from the Earth to the Sun, it was really Edmond Halley who launched the campaigns for observation of the transits of Venus in the years 1761 and 1769, a good example of one of the first campaigns to take place in cooperation between European scientists. Hundreds of observers were sent to different places with the aim of ensuring the success of the campaign, and guaranteeing that there would be good atmospheric conditions for some of the observations. In order to minimise errors, the observers were situated in places whose latitudes were as distant as possible. In the 18th century, travelling to faraway places involved a certain amount of risk, as travellers faced many hazards, on top of which there was open warfare between England and France, especially in the Indian Ocean. The fact is that many of the scientists either died in the attempt or were not able to accomplish the invaluable observations for various reasons.

At that historic moment in time, the great interest in determining the Earth-Sun distance was due to the fact that, thanks to Kepler's third law, they now had a

relationship between all the distances of the planets in the Solar System. Discovering the distance of one of them to the Sun would be enough to open the door to being able to find out the dimensions of our planetary system. Although Halley died in 1742, the scientific community continued with his project. That was how, in 1761, more than 120 observers were coordinated in 62 places and, in 1769, there were more than 151 observers in 77 different locations. A range of dramatic incidents took place, and the results did not all live up to expectations. In both campaigns, the difficulties mainly stemmed from problems on the journey itself, the actual positions of the observers (not always well chosen, particularly regarding longitude) and the clocks they were equipped with being unreliable. Nevertheless, the 1769 expedition was able to count on the experiences of the previous one, which enabled certain factors to be improved. One of them was the 'black drop' effect, observed for the first time in 1761. This effect is due to different causes, one of them being the existence of atmosphere on Venus. The better the resolution of the instrument used, the less this effect appears, but it always does appear, due to the decrease in brightness of the solar surface in the proximity of the solar limb, which causes the observer to make an error when determining the length of time that the internal contact takes place; the uncertainty regarding the actual moment of the contact can be between 20 seconds and one minute. In the past, some expeditions even constructed artificial models for each observer to determine by experiments the extent of 'their error' caused by this effect and thus be able to provide a better reading of the length of time of the contact.



The black drop effect appears like a drop of liquid stretching towards the edge of the solar limb (the apparent edge of the Sun).

Some years before the transits, Joseph-Nicolas Delisle simplified Halley's method by setting out that it would suffice to observe the instant of Venus' ingress into the solar disc and its exit, and he kept up lively correspondence with other astronomers to prepare for the observation. There were many who launched campaigns to gather

funds to finance their expeditions. The situation was made even more difficult as France and Great Britain were opponents in the Seven Years War, which complicated travel plans somewhat. In fact, there were many astronomers of one or the other nationality who were taken prisoner by enemy forces. We shall not go into the details of all the expeditions that were dispatched from different European countries, but we shall briefly mention some. The French Academy of Sciences organised four expeditions for the 1761 transit, not all of them of the same magnitude. Cassini, for instance, travelled to the Observatory of the Society of Jesus in Vienna and carried out observations together with Archduke Joseph of Austria. Alexandre Guy Pingré, however, journeyed to Rodrigues Island in the Indian Ocean. Shortly after rounding the Cape of Good Hope, in southern Africa, they spotted English ships, which they managed to avoid. A little while afterwards, they came across a French ship which had not been so lucky and had been attacked by the English. Consequently, they were obliged to go to its aid, in spite of the astronomer's protests. Pingré thus lost a lot of time and arrived at his destination with only nine days to spare before the transit. In the end, however, bad weather caused him to miss the beginning and end of the transit and he was able to take measurements only when the clouds allowed him. On top of that, before they could leave the island, it was pillaged by the English, and Pingré was confined there for three months until being rescued by a French ship. On the return voyage, his ship was again captured and he had to disembark in Lisbon, from where he travelled overland to Paris, arriving back (with no results) one year and four months after he had left. But much worse was the story of Guillaume Le Gentil, so much so that it is worthy of its own section (see the box on the following page).

London's Royal Society funded three voyages: one to the island of Santa Helena off the south-west coast of Africa, another to Newfoundland, and a third to Bengkulu, on the island of Sumatra. Hardly had this last one sailed, when it came across a French ship. After a battle, the British ship was left badly damaged and had to return to harbour. On the second attempt, the enterprise had to be cancelled when they arrived at the Cape of Good Hope as Bengkulu had been taken by the French.

Spain carried out several observations from the Imperial College in Madrid and from the Spanish Navy Observatory in Cadiz. The total number of observations was 120, which did not show unanimous results, as the astronomers who analysed them obtained, for the solar parallax, values that ranged from 8"28 to 10"60, partly due to the problem mentioned above, the black drop effect, and also to the uncertainty in determining the geographical longitude of the observation points.

THE UNFORTUNATE EXPEDITIONS OF GUILLAUME JOSEPH HYACINTHE JEAN-BAPTISTE LE GENTIL DE LA GALAISIÈRE

For Venus's 1761 and 1769 transits, Le Gentil (as he is known) took part in the two observations organised by the French Academy of Sciences. For the first one, he planned to travel to Pondicherry, a French colony in the south-east of India. Le Gentil's expedition sailed from Brest on 27 March 1760, with plenty of time to reach and prepare the site properly. After suffering several setbacks due to the war between France and England and having to put up with bad weather, and even damage caused by a hurricane, and when very near to the end of the voyage he discovered that Pondicherry had been taken by the English and so had to set sail for home. Le Gentil was only able to carry out his observation at sea, which was of no use whatsoever, as he did not know his coordinates. Disappointed by his failure, he decided to stay in the area to observe the new transit from Pondicherry, where, this time, he arrived 14 months early. But, once more, luck was not on his side, as clouds prevented him from seeing the second transit. Le Gentil arrived back in France in 1771, after an absence of 11 years, 6 months and 13 days, only to find that he had been declared dead and his heirs were dividing up his property and possessions. It took him years, wealth and a great deal of effort to recover them. Clearly, Le Gentil was not blessed with good fortune. As he himself wrote in his story of the voyage:

> "This is the fate that often awaits astronomers. I have journeyed nearly ten thousand leagues; I seem to have crossed such a great expanse of seas, exiling myself from my homeland, only to be the spectator of a fatal cloud, which came to present itself in front of the Sun just at the moment of my observation, and take with it the fruit of my hardship and my fatigue."

The astronomical community threw all their support behind the 1769 transit in the hope that the outcome would be better than that of 1761, and it was. The English organised three expeditions, two of which are mentioned in the Appendix. The French also organised three, one led by Le Gentil, who again had numerous problems (see panel, above); another by Pingré, who went to Santo Domingo and this time suffered no major incidents; and for the third one, Abbé Chappe travelled to California accompanied by two Spanish naval officers. Both the English and the French requested permission from Spain to carry out observations in American territory, as had previous expeditions by the Royal Society and the French Academy

of Sciences when they carried out geodesic measurements in American territories with the aim of determining the shape of the Earth. The scholar and naval officer Jorge Juan, who had been part of the geodesic expedition, gave his opinion and some clear recommendations to the Spanish government:

> "These gentlemen's task consists of doing everything they possibly can: there is no port, fortification, route, town or desert that they do not want to examine, make maps of and make public the most specific details of it all. This is in no way a good thing..."

Consequently, Spain only acquiesced to cooperating in the expedition led by Jean-Baptiste Chappe, whose accompanying pair of Spanish naval officers were equipped with a set of instruments so that they could carry out observations independently of the French group. They sailed from Cadiz on the 21 December 1768, and, after crossing the Atlantic and Mexican territory, arrived at the Pacific coast on 15 April. They set out for California, but had such bad luck, with either the wind against them or no wind at all, that it was 18 May by the time they finally spotted the Californian coast. As the transit was on 3 June, Chappe insisted on disembarking on the coast close to the mission of San José del Cabo, even though the area was subject to an outbreak of typhus. The fear of missing the observation prevailed over all other options and they managed to carry out a good observation. Chappe and Salvador Medina, however, fell victim to typhus and perished, as did a large part of the expedition. It should be added that, in relation to the Spanish, the transit was also observed by groups of scientists in Cadiz (Spain), in Mexico, and in Santa Ana (California).

In all, 151 observers followed the transit from 77 different observation points. The final results, after analysing the whole of the observations, varied from 8"43 to 8"80, which was not bad considering the perturbing effect of the black drop. In fact, in the 19th century, with better methods for processing data and correcting the coordinates of the stations, and the same data, Simon Newcomb obtained a parallax of 8"79, a value which is very near to the one accepted nowadays.

The 19th-century transits took place in 1874 and 1882. Now the astronomers' interest was not just focused on determining distances within the Solar System, they also hoped to determine the distance to the nearest stars. As mentioned before, in 1838 Friedrich Wilhelm Bessel had managed to measure the first stellar parallax, that of star 61 Cygni, and at the end of the century the parallaxes of another score

of stars had been measured by using as a base the distance between two opposite points on the terrestrial orbit and by carrying out observations at intervals of six months. It was, therefore, very important to determine the solar parallax with the highest possible accuracy.

In these transit observations, it was hoped that photography would solve the problem of the black drop by determining the exact instant of the ingress and exit of Venus, but it was not to be. Not even with the help of photography was it possible to eliminate the effects of the atmospheres of Venus and the Earth. In any case, the results obtained in 1874 were good, as they were between 8"79 and 8"83. For that reason, the 1882 transit was observed with less expectation – to be able to significantly increase the number of decimal places, other methods would have to be used.

It must be added that, nowadays, the transits do not have the same importance for determining distance as in the past, but it is interesting to note that this same concept is still used in certain research projects to detect extrasolar planets.

The transits of extrasolar planets

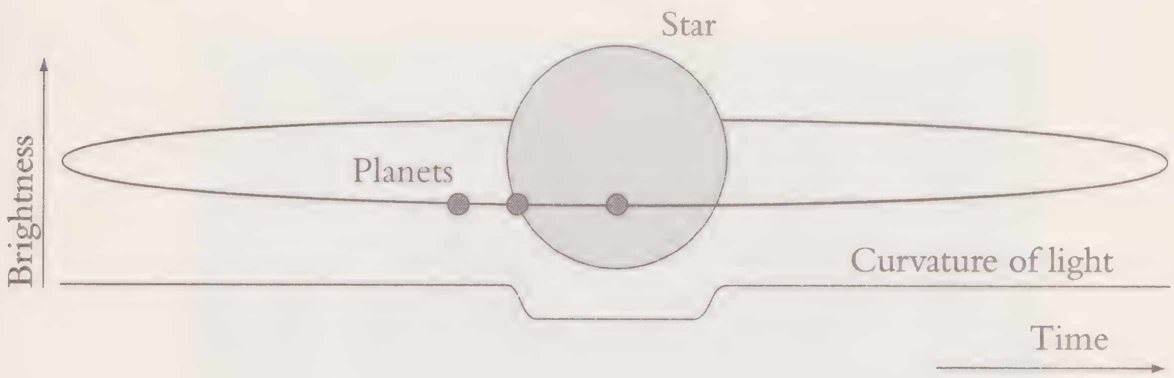
Nowadays the observation of planetary transits is one of the methods used for detecting planets in systems other than our own. On 6 October 1995, Michel Mayor and Didier Queloz, of the Geneva Observatory, announced the discovery of the first exoplanet, 51 Pegasi b. In the search for planets similar to ours, on which there could be analogous conditions, the discovery of the first exoplanet did not live up to expectations. Its orbital period was 4.2 days, and its mass was approximately half the mass of Jupiter. The new planet was "a very hot Jupiter revolving around its star within the orbit of Mercury".

Consequently, the temperature and climate must be extreme and, if it was a gaseous planet, a short life could be forecast for it. Since then, hundreds of exoplanets have been discovered in hundreds of different systems, and they are distributed in all parts of the Universe. Most planets are quite massive and it might appear that the planetary systems do not seem to be very much like ours, though this may be because the methods used to make these discoveries favour the detection of the most massive planets. Four methods are used:

1) Taking direct images of the planet and of the star.
2) Observation of its own motion induced by the planet.

3) Variation in the star's radial velocity due to the presence of the planet.
4) Variation in the star's brightness due to the planet's transit in front of it.

The first three methods can be applied to the case of quite massive planets, such as Jupiter, but the last one, the one using transits, can be used in the detection of planets similar in size to the Earth. When a planet passes in front of a star, the latter's brightness decreases, for which reason the transit can be clearly detected in the curvature of its light. It is a method that is even within the reach of amateur astronomers, though the main problem is that it is only applicable to those planets whose orbital plane is close to the Earth's orbital plane, which is the same as in the case of Venus's transit, which is only observable when Venus and the Earth are near to the nodal line. On top of that, it is only possible to observe this variation in the curvature of the light during a small part of its orbital period, which may only be a few hours. This means that the probability of discovering a planet of a type analogous to the Earth is drastically reduced, but that won't stop us looking.



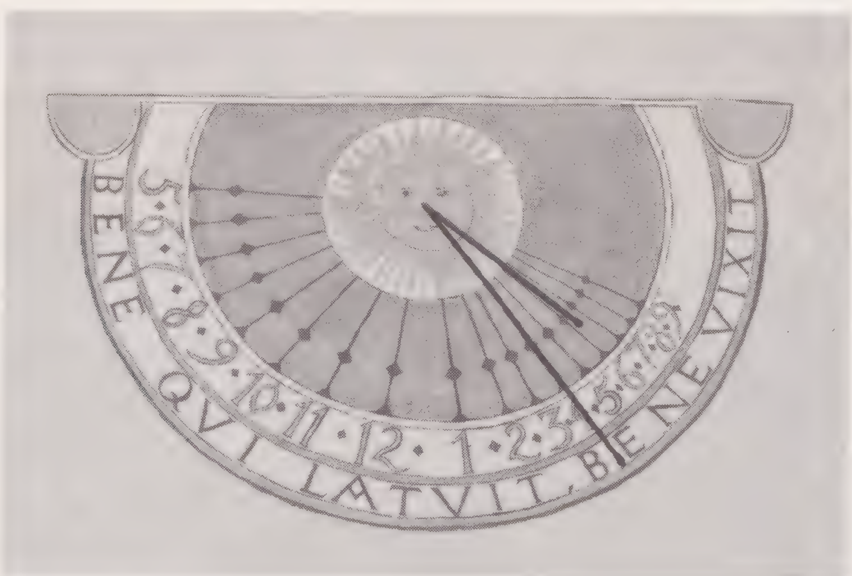
Changes in the path of the light from a star due to an exoplanet passing in front of it.

The probability of observing a planet situated at a distance of one astronomical unit (AU) is 0.5 per cent. In other words, if all the stars have a planet at one AU, it is necessary to follow 200 stars to see a transit. If 10 per cent of the stars have a planet at one AU, we have to watch 10,000 stars to detect five planets.

Chapter 4

Measuring Time

Since its beginnings, an important part of astronomy has been concerned with the measurement of time. When primitive peoples took note of regularity in the changes in weather due to the seasons, they attempted to predict the best times for sowing and collecting their crops, and this led them to acquire their first knowledge of astronomy. Day and night and the seasons prompted man to make the first sundials and calendars, which are devices with a decisive mathematical component. Our aim here is not to compose a treatise on these instruments, but we do intend to show the main guidelines for building a sundial and, what is yet more interesting, to provide an understanding of its basic concepts.



Sundials show the Sun's real time wherever they are set up. This time does not coincide with the time shown on our watches.

The great tragedy of sundials: you have to know how to tell the time

Sundials show the solar time, which is not the same as the time displayed on watches and clocks. That is the problem. Anyone who looks at a sundial, if they don't know

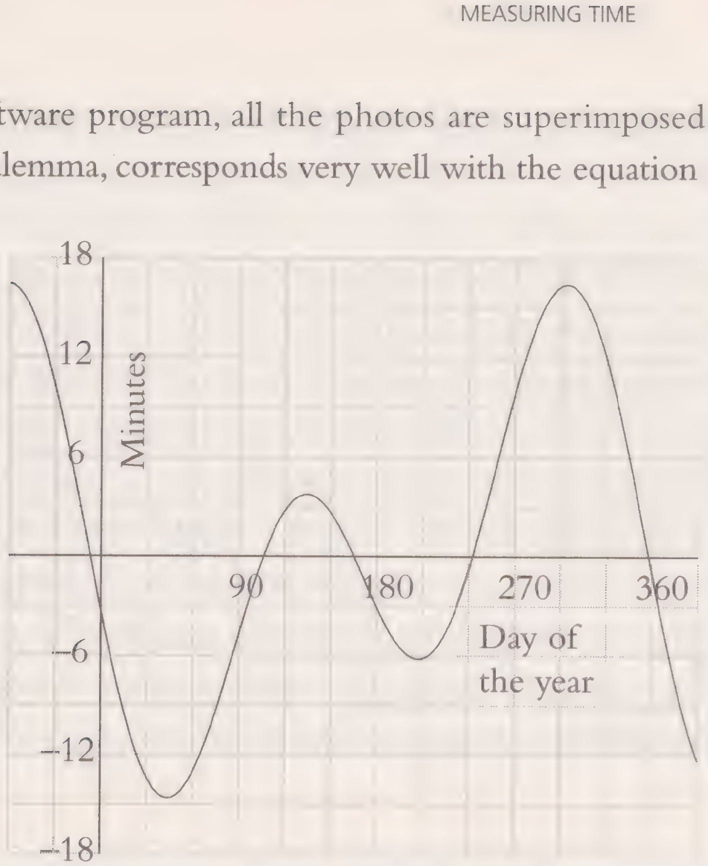
how to interpret the time, will think that it isn't working properly. The problem is that solar clocks show the time at that location according to the sun, and that is not what we nowadays think of as 'the time'.

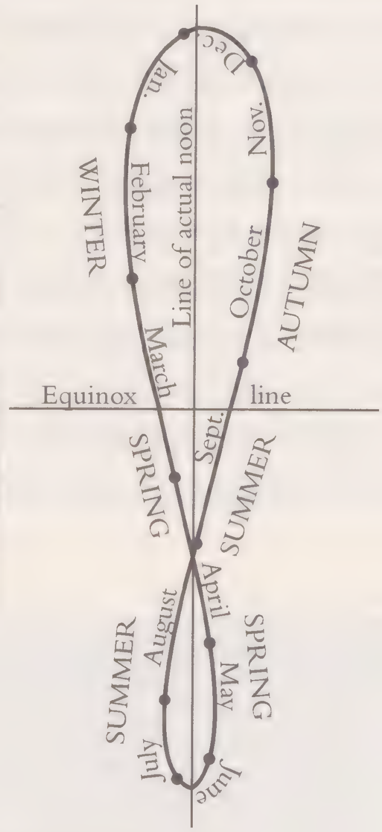
Before man invented mechanical clocks, 'the time' was defined by solar time, and it was the time marked by sundials. But the Earth, revolving around the Sun, follows the Law of Areas and Kepler's second law, which that states the star does not have a constant speed. Consequently, the relative movement of the Sun as seen from the Earth is not always the same. This was a great problem when it came to making a mechanical clock. It was not an easy task to build a device that marked off hours of a different length depending on the time of the year; so, the simplest solution to the problem was to 'define' a 'fictitious Sun' that would count the same length of time in a year as the real Sun, but would do it at a constant speed. The difference between the two suns is tabulated in what is called the 'equation of time'. As can be seen in the table, the difference positive or negative is, at the most, a quarter of an hour, but this tends to make the observer think that the sundial is not working correctly because it does not coincide with the time shown on their wristwatch.

Equation of time (figures in minutes)												
Days	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
1	+3.4	+13.6	+12.5	+4.1	–2.9	–2.4	+3.6	+6.3	+0.2	–10.1	–16.4	–11.2
6	+5.7	+5.1	+11.2	+2.6	–3.4	–1.6	+4.5	+5.9	–1.5	–11.7	–16.4	–9.2
11	+7.8	+7.3	+10.2	+1.2	–3.7	–0.6	+5.3	+5.2	–3.2	–13.1	–16.0	–7.0
16	+9.7	+9.2	+8.9	–0.1	–3.8	+0.4	+5.9	+4.3	–4.9	–14.3	–15.3	–4.6
21	+11.2	+13.8	+7.4	–1.2	–3.6	+1.5	+6.3	+3.2	–6.7	–15.3	–14.3	–2.2
26	+12.5	+13.1	+5.9	–2.2	–3.2	+2.6	+6.4	+1.9	–8.5	–15.9	–12.9	+0.3
31	+13.4		+4.4		–2.5		+6.3	+0.5		–16.3		+2.8

The graph on page 99 shows the equation of time by month. Sometimes it folds over itself to form a figure-of-eight, or analemma. In reality, this latter aspect is the one that better coincides with the visualisation of this phenomena in the sky. If photographs of the Sun are taken regularly over the same horizon on different days of the year at the same time (the time on a mechanical clock, that is, without the correction for the equation of time having been applied) and, by using a suitable

software program, all the photos are superimposed, the resulting final image, an analemma, corresponds very well with the equation of time.





Above: the equation of time through the year. Right: a representation of the equation of time, in the form of an analemma, taking the location's true noon as a reference.

The displacement of the real Sun in respect to the fictitious one is revealed by a series of superimposed photographs all taken at the same 'clock time'.

That is not the whole story, however, as the subject is still more complicated. This is because the sundial gives the location's real time, while the wristwatch does not show the 'real' time but instead works with a built-in time defined for the whole of the country or, if it is large country, for the whole of a zone within it, or, in the case of a small country, its time is coordinated with other countries. For example, practically the whole of western Europe goes by the time of the Sun passing over the Greenwich meridian. It is convenient to be able to travel around Europe without having to be continually changing the time; our time is, therefore, decided by agreement. If we want to compare it with the real time on a sundial, which is the local time, the longitude has to be taken into account. The world is divided into 24 time zones starting from the prime meridian, ie, the Greenwich meridian. To make the adjustment for longitude, the local longitude and the longitude of the 'standard' meridian of the zone need to be known; a + sign is added to the east and a – sign

DUODECIMAL AND SEXAGESIMAL BASES IN ASTRONOMY

Thanks to Mesopotamia, Egypt and Greece, present-day society, though it uses a decimal system for nearly all applications, still sees a day as being 24 hours, an hour as being 60 minutes and a minute as 60 seconds. Babylonian mathematicians used a numeration system that was sexagesimal (base 60), from which the division of units into 60s is derived, and which has been maintained in astronomy for measuring angles, coordinates and time. The Babylonians' progress in mathematics was perhaps helped by the fact that the number 60 has a lot of divisors (1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30 and 60), making calculation with fractions easier. The Babylonians had a true system of positional numbering, in which the digits written to the left represented values of a higher order, as in our modern decimal system of numeration. They lacked, however, any equivalent to our decimal point, and so the ability to use fractions was of great importance to them.

Most historians believe that the Egyptians were the first to divide the day into smaller parts thanks to the development, around 1500 BC, of a T-shaped sundial which marked off the day in 12 intervals of time. This division reflected the usage of the duodecimal system (base 12) used by that culture. Both the day and the night were divided into 12 parts, which gave rise to the concept of a day of 24 hours, although these hours were of a different length of time. However, to arrive at the concept of fixed hours, it would be necessary to wait until Hipparchus

towards the west. The longitudes have to be expressed in hours, minutes and seconds (1 degree = 4 minutes of time).

But, on top of that, for some years now, many countries, to reduce the consumption of energy for economic reasons, have a summer and a winter time. In summer, they usually put the clock's forward by an hour. The summer/winter change to the clocks is a decision for the government of each separate country, and the change is usually made on Sundays from two to three o' clock in the morning. So if we take into account the equation of time, the correction for longitude and the summer/winter change, the chances that a sundial will show the same time as the watch on our wrist are practically zero. On top of that, we have to take into account the good intentions of people who, while not having much knowledge of astronomy, decide to repaint the dial and 'reset' it. They think that the solution is simple. In the same way that a mechanical clock is set by moving the minute hand until it shows the correct time, they try to do

of Nicaea (c.190 BC–c.120 BC), who was the first to introduce the division of the day into 24 equinoctial hours (based on the 12 hours of light and 12 hours of darkness on the equinoxes). Despite his proposal, for many centuries the length of time of an hour varied according to the season. In fact, hours of a fixed duration were not standardised throughout the world until the appearance of the first mechanical clocks in Europe.

Hipparchus and other Greek astronomers used a variety of techniques previously developed by the Babylonians and Sumerians. Eratosthenes (276 BC–195 BC) divided the circle into 60 equal parts to build the first system of horizontal lines that corresponded to places on the same latitude. Hipparchus completed the system one century later with a system of parallel longitude lines covering 360 degrees. It was not until 150 AD that Claudius Ptolemy, in his book *Almagest*, a continuation of his previous work, made a subdivision of each of the 360 degrees of latitude and longitude into smaller parts. The first division, the *minutiae primae* parts, became known simply as the minute; the second division, the *minutiae secundae*, was shortened to second. The minutes and seconds, however, were not used until several centuries after *Almagest*. Clocks were divided into halves, thirds, quarters, and sometimes 12 parts, but never into 60. In general, people did not use minutes until the first mechanical clocks appeared, which was towards the end of the 16th century.

the same with a sundial and move the gnomon so that it points to the right time. With this drastic solution, the sundial has no chance of showing the correct time.

The writer of this book has a collection of photos of sundials that have been suspiciously 'repaired' which could be used to illustrate each of the adjustments that need to be done, as explained above. I would also like to add an anecdote to this explanation. Some years ago, I had the opportunity to visit a country house that had recently been restored and had a very well restored sundial. I congratulated the owner on the quality of the sundial restoration, but he answered: "You must be joking. It doesn't work. We even called in a lecturer from the university who came to fix it. But he came at night. Obviously, how can it be expected to work right if he came to fix it in the dark?"

In actual fact, the sundial was working perfectly well, and the university professor went in the dark so as to be able to check that the gnomon pointed to the celestial pole, the Pole Star, which is just how he should have done it. The lines were well plotted and the errors were due to adjustments needed for the equation of time, longitude, and for summer time, which it was when I sent to see it. I tried to explain all this to the owner, and he listened to me very politely, but I'm not sure if I managed to convince him or whether he was thinking: "These university folk always stick up for each other."

Be that as it may, sundials form part of our cultural heritage and we shall give them a brief review so as to provide an understanding of how they work. Additionally, those interested may even decide to build one.

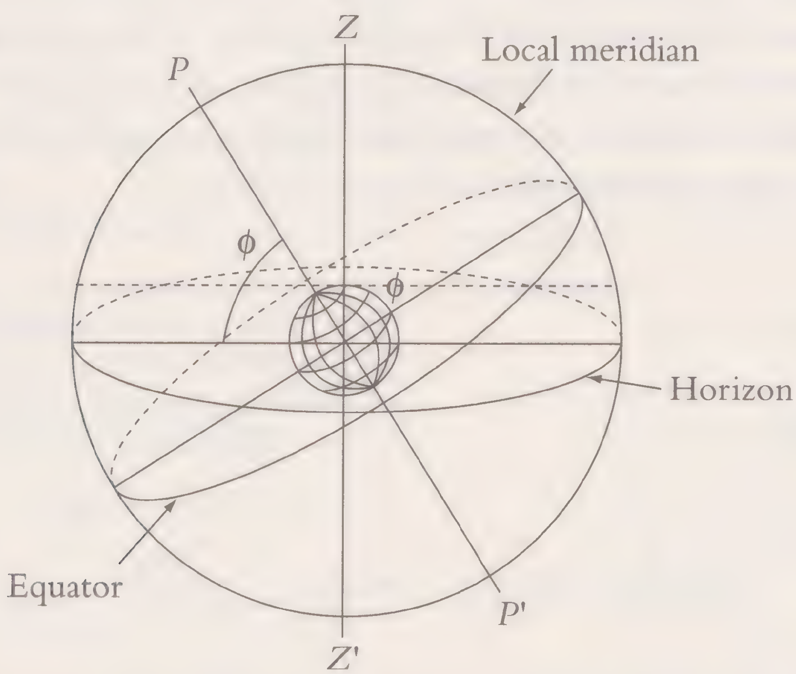
Sundials for measuring time

The fundamental parts of a sundial are its gnomon and its plane. Depending on this plane, the dial can be classed as equatorial, horizontal or vertical. There are, of course, other more elaborate types of solar clocks, but we shall not be covering them in this book. Our aim is not to provide an exhaustive treatise on solar clocks but to give a reasoned explanation of the underlying astronomical concepts and the mathematical concepts behind them.

Sundials are based on the Sun's observable relative movement with respect to the Earth. Because of the planet's rotation on its axis from west to east, we get the impression that the Sun rises in the east and sets in the west every day. As we see the Sun move in relation to the Earth's axis of rotation, the dial's gnomon has to be oriented in accordance with that axis of rotation, regardless of where we install it.

So, it is important to know the position of the place where we are going to build the clock, and in particular we need to know the latitude. (The longitude is only necessary for reading the hour, but we shall look at that later).

So that the gnomon is pointed in accordance with the Earth's axis of rotation, we must orientate it towards the North Star, that is, the celestial north pole, provided that we are in the northern hemisphere, and towards the celestial south pole if we are in that hemisphere. In either case, the angle formed by the gnomon with the plane of the horizon must be equal to the latitude of the place where we are building the clock.

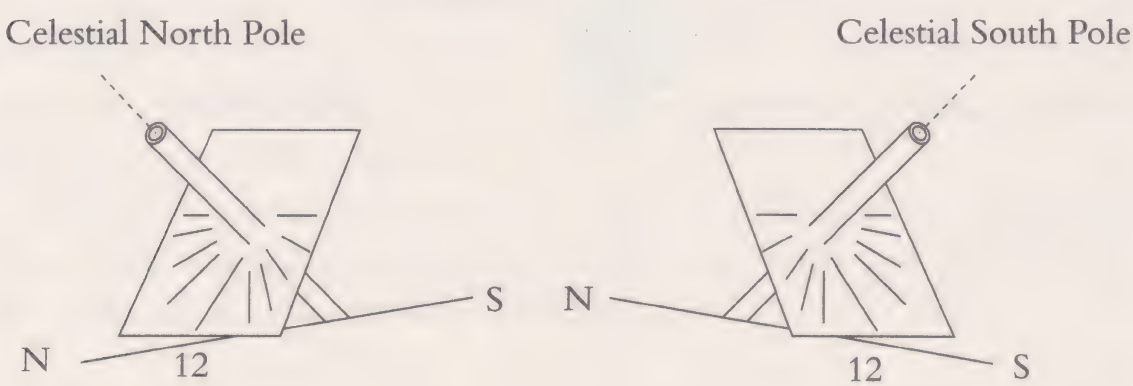


The height of the celestial pole over your horizon coincides with your latitude on Earth. Note that this drawing is not to any scale, as the radius of the Earth is practically zero compared to the infinite radius of the celestial sphere, but here it was decided to 'cheat' any notion of scale so as to be able to establish the principle by visual means.

As can be seen in the illustration above, the height of the celestial pole over the observer's horizon equals an angle that is identical to the latitude of the observer's location, or angle on the meridian of the place from the terrestrial equator to the position that we occupy on the terrestrial sphere, as the latitude is determined by the plane of the terrestrial equator and the plumb line, or, what comes to the same thing, the height of the pole – that is, from the axis of the terrestrial rotation, with respect to the plane of the horizon. These angles are equal, as they are determined by perpendicular elements between them.

Equatorial sundials

Sundials are classified according to the plane used. We shall begin with the simplest: the sundial with a plane parallel to the Equator. As the Sun follows the celestial equator on the first day of each of the equinoxes (the vernal and the autumnal) and the other days we see it following a parallel above or below until it reaches, at most, the Tropic of Cancer (with a declination of +23.5°) and of Capricorn (with a declination of –23.5°), the simplest way to construct a solar clock is to have the plane of the clock parallel to the plane of the Equator, and the gnomon in accordance with the axis of terrestrial rotation, as shown in the diagrams below. Thus we will have a gnomon which forms an angle with the horizontal that is identical to the latitude of the location and oriented towards the celestial pole. This is why the gnomon has to be situated on the north-south line. To do this we can make use of a compass and assume that there will really be a small error due to the magnetic pole declination, which continually varies, but we can overlook it.



The alignment of an equatorial sundial in the northern hemisphere (above left) and in the southern hemisphere (above right).

The plane of the Equator will be set perpendicular to the axis of rotation and, therefore, it will also be placed perpendicular on the north-south line on the plane of the horizon. The line, on the dial plane, from the point of intersection of the gnomon to the point of intersection of the plane with the north-south line on which the dial is placed, is the hour line for noon, 12:00. Obviously, it will be the solar noon when the Sun passes over the north-south line; the hours before will be morning and the ones after will be afternoon. We place all the remaining hours at equal distances of 15°, as the Sun makes a complete revolution of 360° in 24 hours.

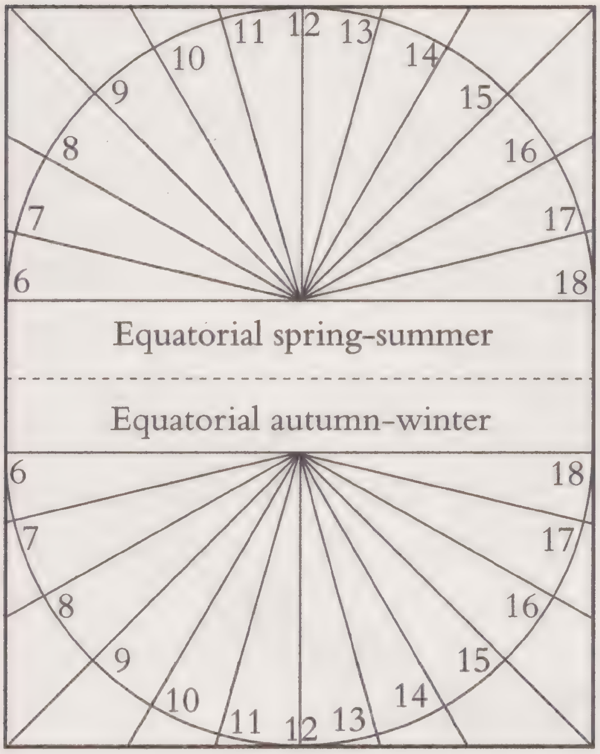
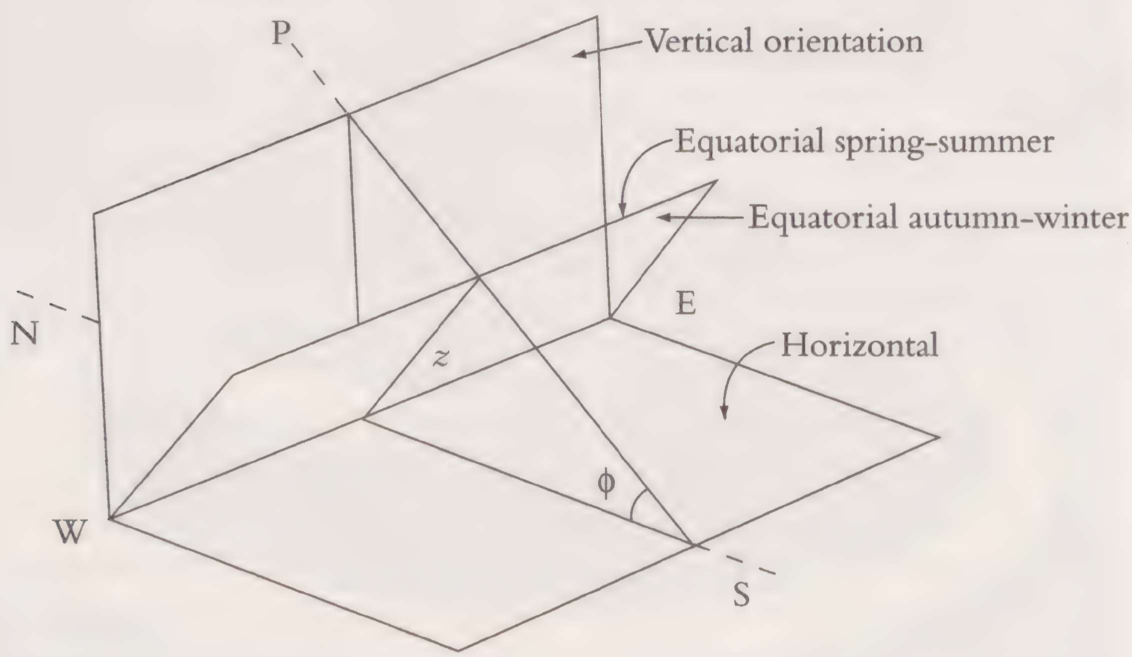
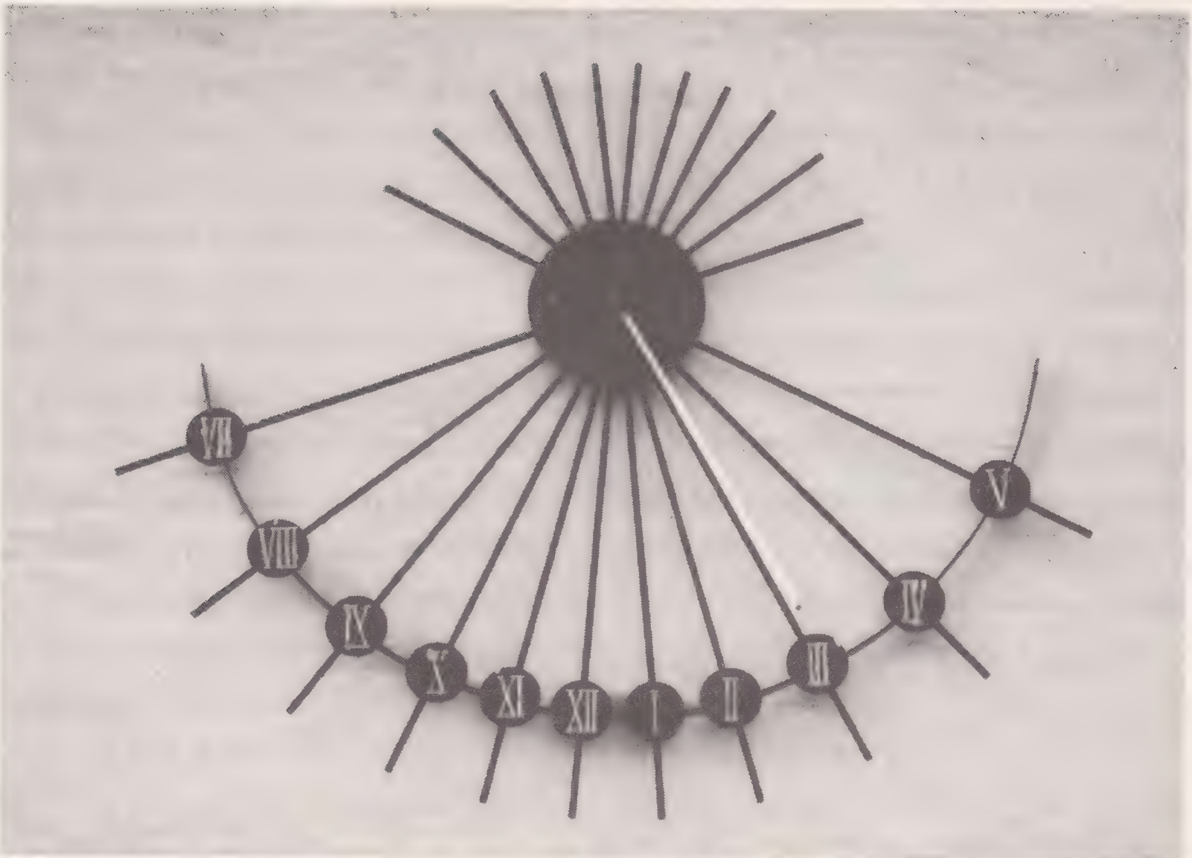


Diagram of an equatorial sundial face with the hours every 15 degrees. It has to be folded along the dotted line as this sundial shows the time with the sun above the equatorial plane for half the year and below it for the other half.

These are certainly the simplest sundials to build, but they have a peculiarity – they show the time with the sun above the equatorial plane in spring and summer and below in autumn and winter; therefore, the lines for the hours have to be drawn on the two sides of the plane, as shown in the diagram (left). The fact is that, although they are the simplest, they are not the most common as the majority of sundials have a horizontal or vertical plane. All of them can be deduced from the equatorial example by carrying out a simple projection and by using trigonometry.



Three sundials in one: equatorial, horizontal and vertically oriented. The figure shown is for the northern hemisphere; for the southern hemisphere the position of the cardinal points have to be flipped.



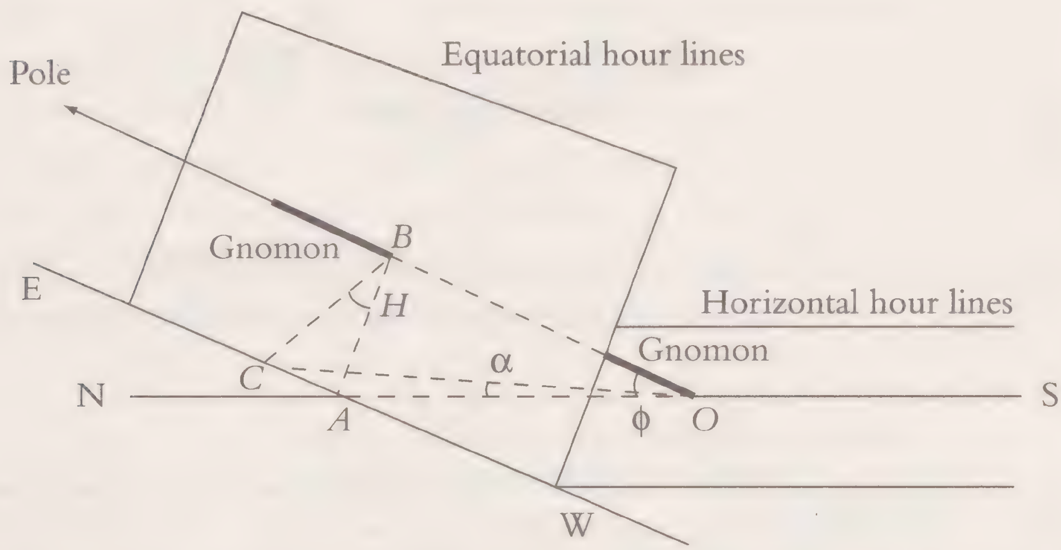
Top: a vertical sundial. Above: a horizontal sundial cast in bronze.

Horizontal sundials

These sundials are on the horizontal plane, which has to be checked with a spirit level to make sure that it is true. The gnomon has to be situated so that it forms an angle equal to the location's latitude, on the north-south line and pointing towards the celestial pole. That line will correspond to the hour-line for 12 noon, and the rest of the hour-lines will have to be calculated by using the expression

$$\tan\alpha = \tan H \sin\phi,$$

with α being the angle between the 12-noon hour-line and the other hour-lines, and $H = 15°, 30°, 45°\ldots$, respectively, in accordance with the following figure:



The hour-lines of this new sundial can be obtained by projecting the hour-lines of the equatorial dial on the horizontal plane just by taking into account that $\tan\alpha = CA/AO$, $\tan H = CA/AB$ *and* $\sin\phi = AB/AO$, *whence it is deduced that* $\tan\alpha = \tan H \sin\phi$, *where* H *can be* 15° *and then* α *gives us the angle for the hour-line of 11:00 and for 13:00 hours, or alternatively* H *can be* 30°, *and then* α *gives us the hour-line for 10:00 and 14:00 hours, and so on up to 6:00 and 18:00.*

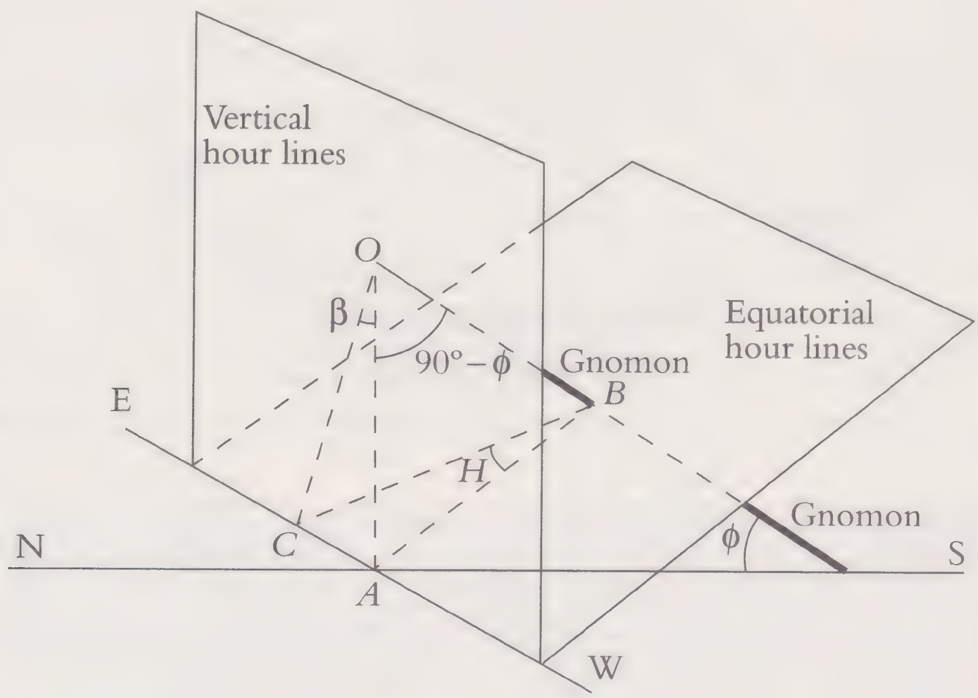
Vertical sundials

Very often sundials are set on the wall of a building, and therefore have the dial plane in vertical position. If the wall is oriented in the east-west direction, they are no more complicated than a horizontal sundial; it just has to be borne in mind that the angle for setting the gnomon on the wall will be determined by the complementary angle of the location's latitude, the colatitude. The hour-lines can also be deduced by

projecting those of the equatorial dial onto the vertical plane oriented in accordance with the east-west line and the expression:

$$\tan \beta = \tan H \cos \phi,$$

with β being the angle between the hour-line of 12.00 and the other hour-line, and $H = 15°, 30°, 45°\ldots$ respectively, as in the next diagram:



The hour-lines of this new dial can be obtained by projecting the hour-lines of the equatorial dial on the vertical plane just by taking into account that $\tan\beta = CA/AO$, $\tan H = CA/AB$ *and* $\sin(90° - \phi) = AB/AO$, *whence it is deduced that* $\tan\beta = \tan H \cos\phi$, *where* H *can be 15 ° and then* β *gives us the angle for the hour-line of 11:00 and 13:00, or alternatively* H *can be 30° and then* β *gives us the line for 10:00 and 14:00, and so on up to 6:00 and 18:00.*

In reality, however, most houses do not have a wall positioned exactly in the east-west direction but, instead, have one somewhat offset with respect to that direction, that is, it will form a certain angle which can be measured exactly. That's where the procedure for plotting the dial's hour-lines becomes quite complicated. The level of trigonometry is higher and appears in the Appendix.

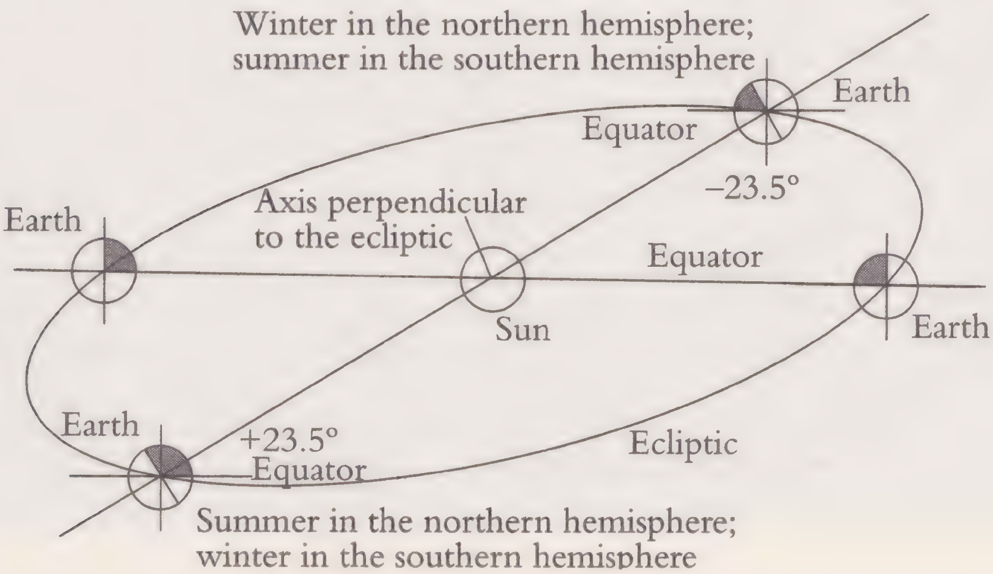
Sundials as calendars

Some solar clocks also serve as calendars and, besides the hour, can also show the approximate day of the month. This is on account of the fact that the Sun does not

move at the same height over the horizon every day, but instead is much higher in summer than in winter. This can enable the month of the year to be read by using the zodiac lines, which in reality are simply conic sections. But to get to these conic sections, let's briefly look at how the Sun moves throughout one year.

The origin of the seasons

The Earth follows translational motion around the Sun on what is called the ecliptic plane. The Earth's rotation axis forms an angle of 23.5° with the axis perpendicular to the ecliptic plane; it is what is called the obliquity of the ecliptic. This inclination of the rotation axis is what gives rise to the seasons, as the Sun as seen from Earth is at a different height depending on the time of the year. And so the Equator forms an angle of, at the most, ±23.5° with the ecliptic plane. The diagram below shows how on midsummer's day in the northern hemisphere, the declination or height of the Sun over the Equator is +23.5°, which coincides with midwinter in the southern hemisphere. And vice-versa – when it is midwinter in the northern hemisphere, which is when the sun is lowest over the horizon as its declination over the Equator is –23.5°, it is also midsummer in the southern hemisphere. On either of the two equinoxes, the declination of the Sun over the Equator is 0°.

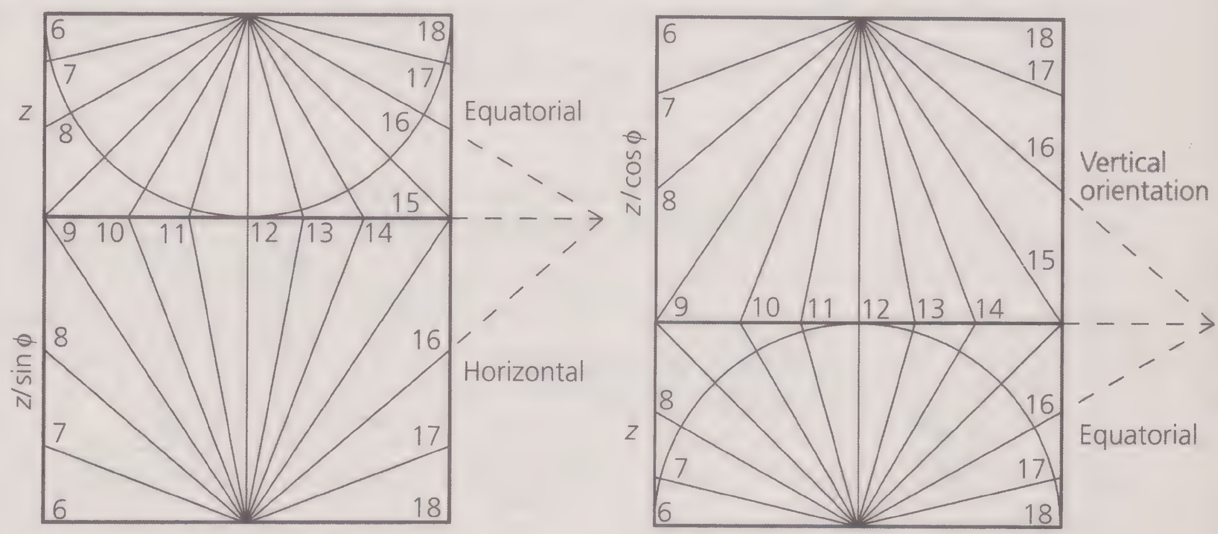


The equator forms an angle of, at the most, ±23.5° in respect to the ecliptic plane; this angle is what gives rise to the seasons.

HOW TO MAKE A POCKET SUNDIAL

The hour lines of a horizontal or vertical oriented sundial are obtained by projection of the equatorial sundial, with nothing else to take into consideration but the location's latitude, as was seen earlier, but now we will show how to make a pocket sundial with both horizontal and vertical faces. We shall begin by looking at the dimensions of the sundial that we are going to build. We shall draw a half-circle of radius z and shall mark the hour-lines from 6 to 18 using an angle of 15°. The dimensions of the horizontal and vertical oriented clocks can be deduced from the figure on page 105, and simply by using the rectangular triangles determined by the rotation axis and the equatorial plane and the definitions of the sine and cosine of latitude. Thus, the 12-line of the horizontal dial will be the distance $z/\sin\phi$, and the 12-line of the vertical one will be distance $z/\cos\phi$.

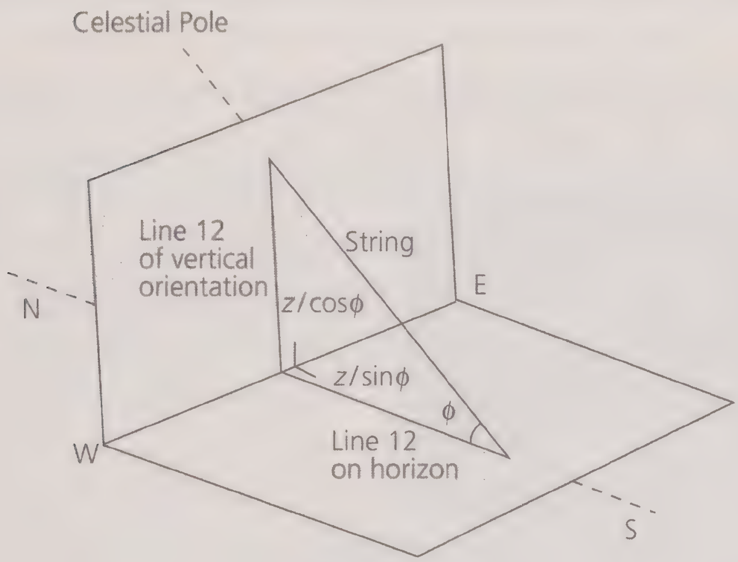
Let's start with the construction of the horizontal sundial. We extend the 12-line and draw on it the distance $z/\sin\phi$ which will give us the dimensions of the horizontal dial. As the hour-lines must be the same on all the sundials, we extend the lines of the equatorial dial to determine their intersection points with the line perpendicular to the 12-line, ie, the north-south line.



Relationship between equatorial and horizontal sundials (above left), and between an equatorial and a vertical oriented sundial.

The height of the Sun over the horizon is important because, the higher it is, the more perpendicular are its rays when they reach the terrestrial surface, meaning that the radiation per square metre is higher; therefore we feel that it is warmer.

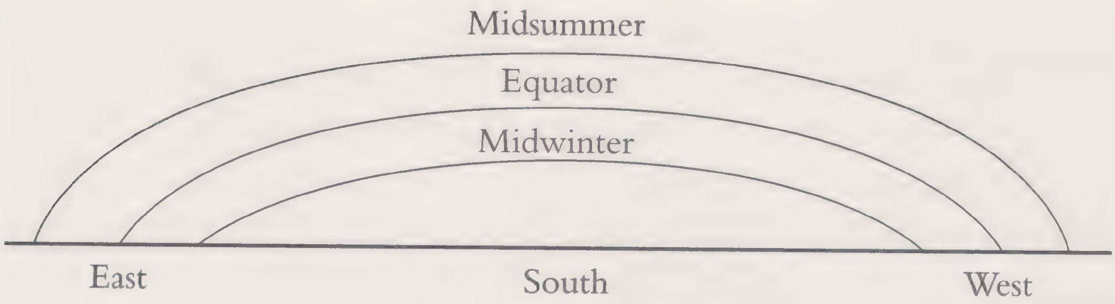
Finally, we copy the planes of the vertical and horizontal dials onto two pieces of wood fixed to each other with adhesive tape (or a pair of good hinges, though in this case the thickness of the hinges has to be taken into account). We shall fasten a piece of string to the two meeting points of the hour-lines so that the 12-hour lines of the horizontal and vertical dials are kept at a right angle. This part is the most difficult and the help of another person is needed to hold the pieces together while you fix the string in place.



When fixed tautly in place, the string acts as the gnomon. It is important that the two dials are perpendicular to each other when the string is fastened.

To be able to use the sundial, the 12-hour line of the horizontal dial must be aligned with the north-south line. For that reason, use a compass so that it can be oriented accurately. It can also be useful to use a spirit level so that the horizontal plane can be placed in a truly horizontal position. You may find a small plumb line helps ensure vertical orientation of the vertical dial. Another possibility is to move the whole set around until you manage to get the two dials showing the same time, but this trial and error approach can be a little difficult.

Do not forget that that the surface area receiving solar radiation in the respective hemisphere is greater in summer than in winter. Furthermore, as the Sun is higher over the horizon, there are more hours of sunshine, and therefore the temperature becomes hotter.



The Sun follows the path of the Equator on only two days a year: the days of the spring and autumn equinoxes, when the length of the day and the night is the same in either of the two hemispheres. In the illustration above, the southern horizon for the northern hemisphere is shown.

URBAN LEGENDS

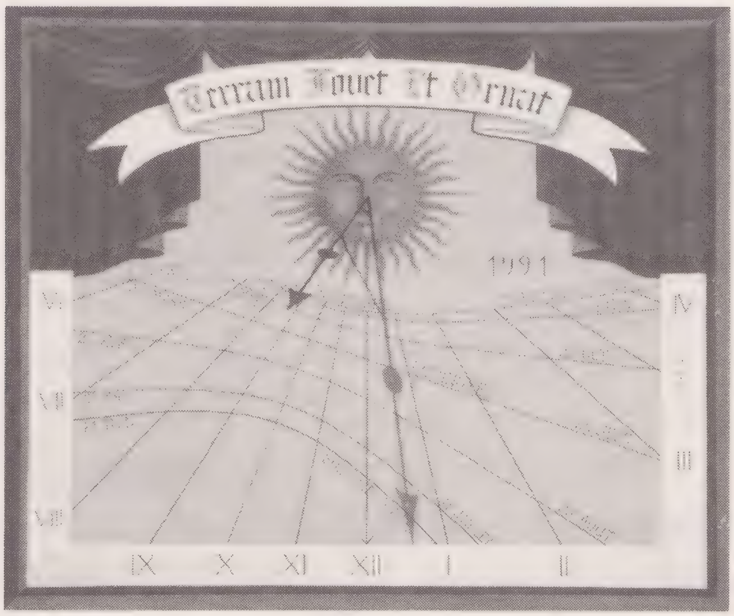
There are a number of theories involving the world of astronomy that we could describe as urban legends, and which should be refuted here. One of them is the belief that the Sun rises in the east. That is not always true: the Sun rises at the east cardinal point and sets at the west cardinal point on two days a year, on the two equinoxes. On the other days of the year, it rises at different points ranging from south-east to north-east, and it sets between south-west and north-west depending on the date.

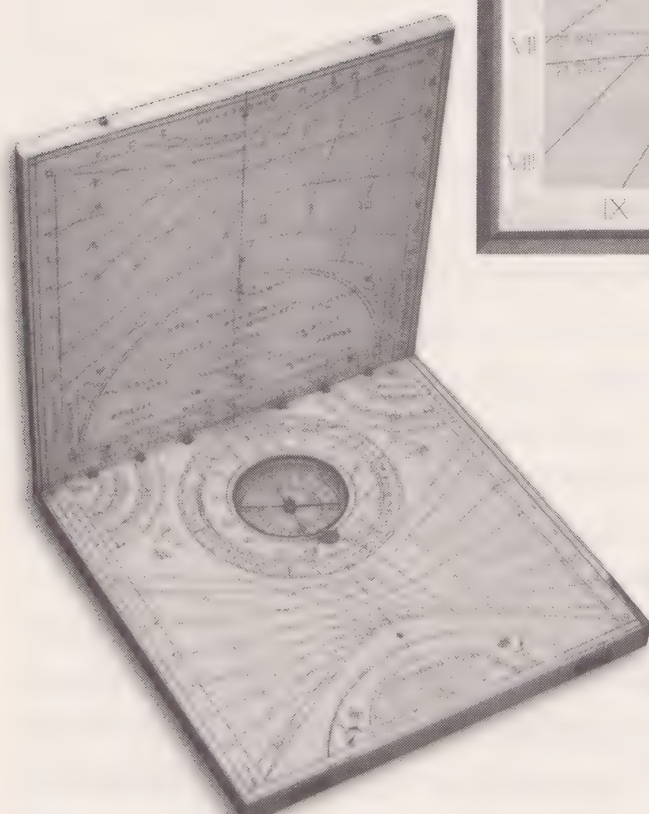
Another popular belief is that in summer it is warmer because the Earth is nearer the Sun. In actual fact, the Earth's orbit around our star is an ellipse with very little eccentricity, hardly any at all. In other words, Earth's orbit of the Sun is almost a circle, so the Earth's distance from the Sun varies little. Furthermore, if the Earth were to move nearer to the Sun at certain times, how could half the planet ever have winter? And, to top it all off, it turns out to be the case that it is summer in the northern hemisphere when the Earth is actually at its furthest from the Sun. As has been explained above, the seasons are a consequence of the tilt in the Earth's rotation axis in respect to its plane of motion around the Sun.

Zodiac lines, tapered paths

As, throughout the year, the Sun changes its height over the horizon depending on its daily declination, this data can be reflected on a sundial. If, for instance, we have a horizontal sundial, the shadow at the end of the gnomon will be shorter in summer than in winter because the Sun is higher in summer. For vertical sundials it will be the opposite: the higher the Sun is, the longer the shadow cast by the gnomon.

There is a line on some sundials that represents the position of the tip of the gnomon's shadow for a certain day. As it is not feasible to draw the line for every day, as the dial would then be completely covered in lines, normally only a line for the first day of each season is shown. Thus, in the northern hemisphere the following are given: a line for 21 March, the first day of spring, which coincides with the passing of the Sun through the Aries point; 12 June, midsummer's day, when the Sun enters into the sign of Cancer; 23 September, the first day of autumn, the autumnal equinox, when the Sun passes through the Libra point, and 21 December, when the Sun enters Capricorn. Note that these dates are approximate, as the passing of the Sun through Aries has to be calculated for each year. For the southern hemisphere, the signs Cancer and Capricorn for the solstices are swapped for winter and summer.



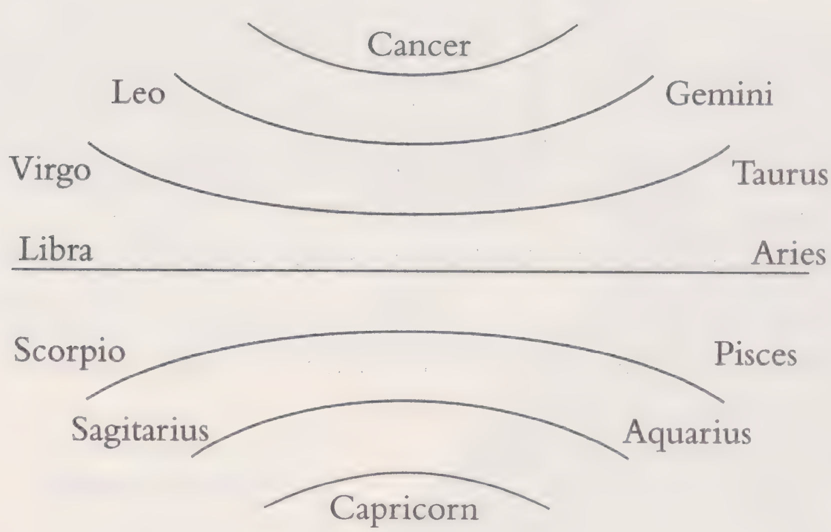


Above: a photograph of a vertical declining sundial on which, in addition to the hour-lines, the zodiac lines are also shown. Left: a pocket sundial with hour and zodiac lines and also a compass for orientation purposes.

There are sundials that not only include the lines mentioned above, but also lines corresponding to the tip of the gnomon's shadow throughout the year for the first day of each of the zodiac signs. In any case, the mathematical process for plotting the lines is the same as that for the four seasons.

The zodiac lines on a sundial do not add up to twelve, nor do the lines corresponding to the seasons come to four, as some of them overlap. The Sun runs along the Equator on the spring (or vernal) equinox; on the day after, the sun runs along a parallel a little above the Equator; the day after that, the sun covers a parallel a little bit more above, and so on, raising the declination day by day until it reaches its maximum of 23.5° at midsummer. On the day after the solstice, the Sun begins its downward path by running along a parallel a little below, and so on until the autumn equinox when it again runs along the Equator.

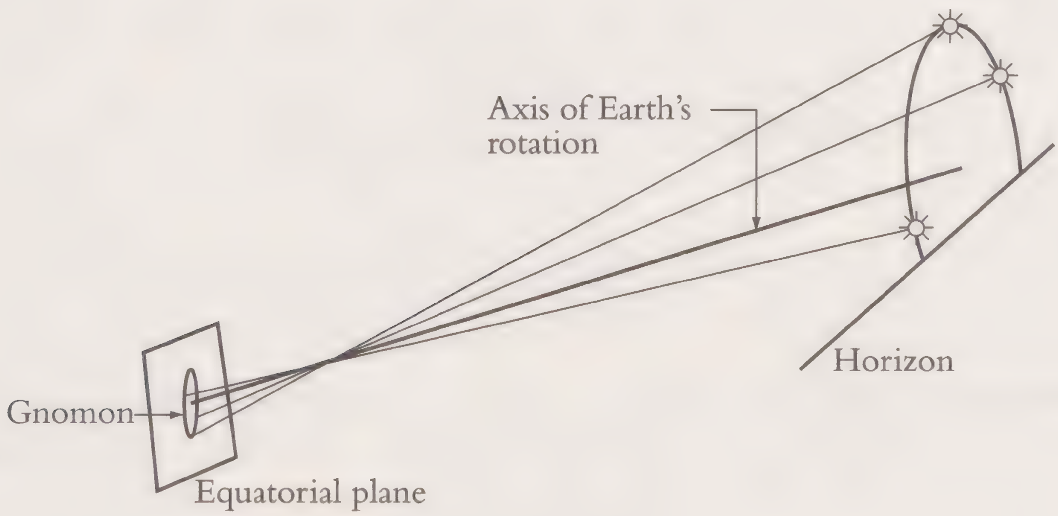
So, we have established that the two seasons 'overlap' in the same way that the two equinox days do; therefore, if we draw the zodiac signs corresponding to spring, they will overlap those for summer, in the same way that the Aries line (spring equinox) coincides with that of Libra (autumn equinox). So, the Taurus line coincides with that of Virgo, and the Gemini line with Leo's. The trajectories of the autumn and winter signs will also coincide: Scorpio with Pisces and Sagittarius with Aquarius, as the diagram below shows:



But the next issue is to ask ourselves why the tapered pattern appears. There is a simple answer: just imagine the daily path of the Sun around the terrestrial rotation axis in parallels. If we imagine that the Sun has just one sun ray that arrives at the tip of the gnomon, as the Sun revolves around the world's rotation axis, that ray would

give rise to a conical surface with its vertex at the end of the gnomon which, let's not forget, is always in the direction of the Earth's rotation axis.

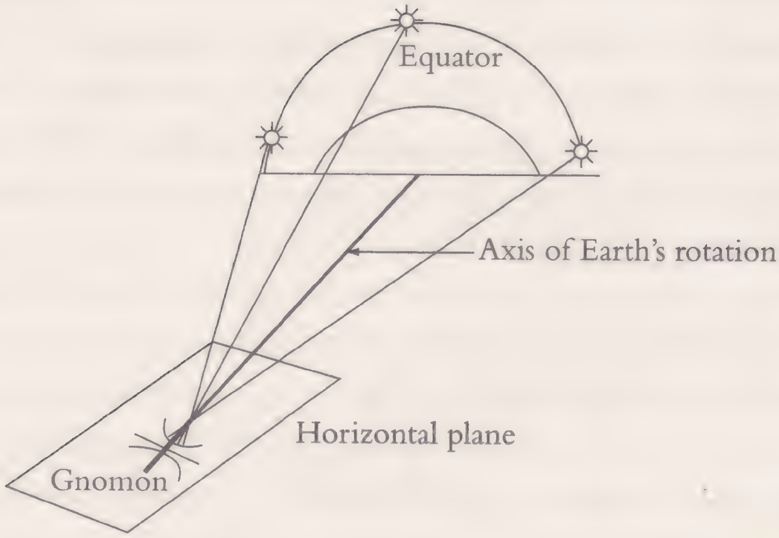
If we cut that cone with a plane parallel to the Equator, in other words, perpendicular to the rotation axis, the intersection will be a circle. The zodiac lines of equatorial sundials are circles, and their radius depends on the Sun's declination and the length of the gnomon.



The zodiac signs of equatorial sundials are concentric circumferences whose centre is the point of intersection of the gnomon and the plane of the sundial.

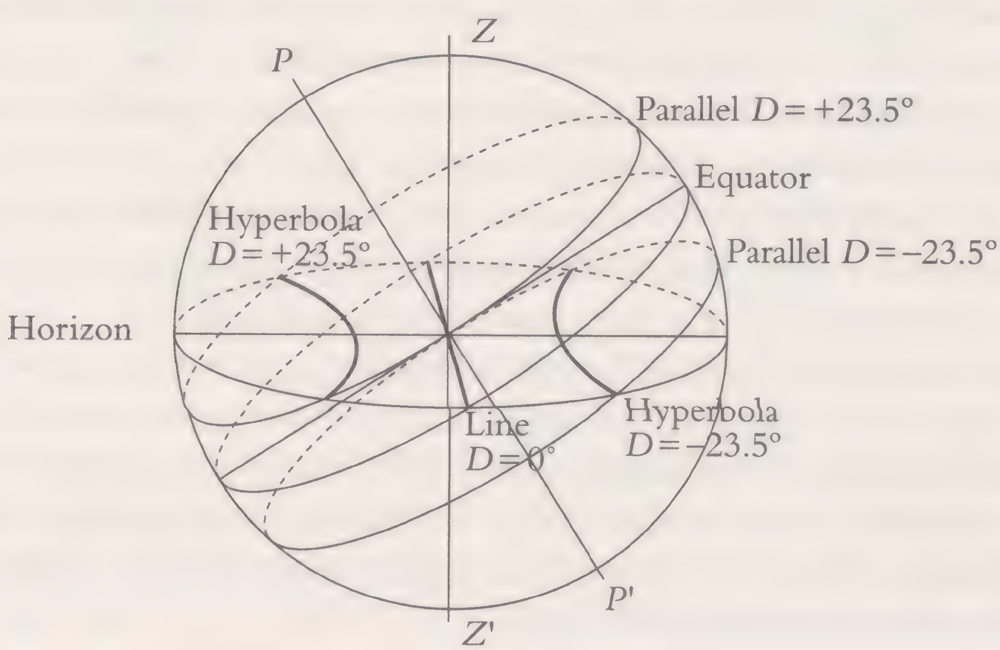
But if we cut through a horizontal or vertical surface, the intersections obtained are branches of hyperbolas, whose shape depends on the location's latitude and, obviously, on the Sun's declination when it enters each zodiac sign. It so happens that, depending on the Sun's declination, the branches of the hyperbola are concave or convex, becoming, by continuity, a straight line right on the day of the equinox. If the celestial sphere is really represented with infinite radius and the Earth is considered just a point, then the representation of the cones on the plane of the horizon is simplified as in the lower diagram on the next page.

It is obvious that if the tip of the gnomon runs along one of the zodiac lines or a line between two of them, we know approximately what day of the month it is. For example, if the tip of the gnomon is between the Aries-Libra line and the Scorpio-Pisces line and the leaves are beginning to fall from the trees, we are in October, but if there are no leaves on the trees, it is February. Before mechanical clocks were invented, this method was good enough.



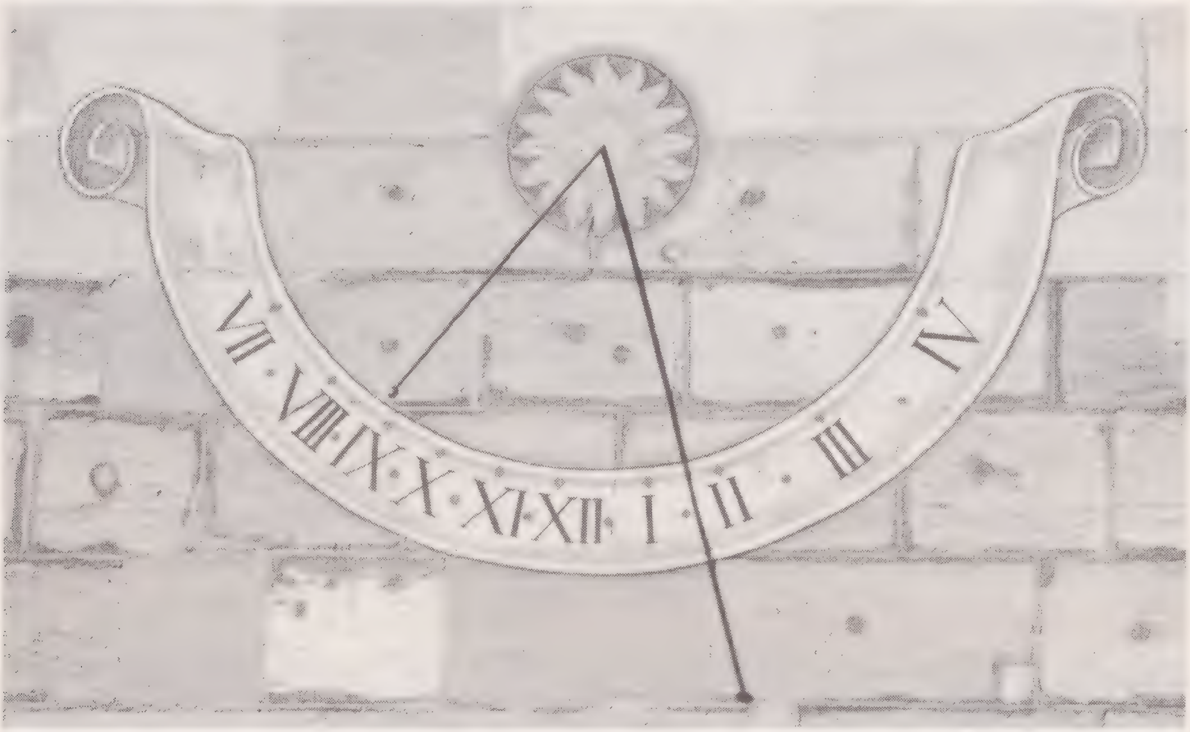
Zodiac lines on horizontal or vertical sundials are hyperbolic except on the equinoxes, when they are straight.

The diagram below shows sections of the conic surface determined by the Sun and the tip of the gnomon on the horizon plane. On the sphere of infinite radius, the Earth is reduced to a point and the gnomon too. The intersection of the cone determined by the parallel of declination $D = +23.5°$ with the horizon plane creates a hyperbola; the intersection with the parallel of declination $D = -23.5°$ creates another hyperbola. For other declinations intermediate hyperbolas are obtained except in the case of declination 0°; in that case the cone is reduced to a circle and the intersection with the horizon plane is a straight line.



How to solve the clock problem

Formerly, sundials were not expected to be very accurate. It seems that in the 17th and 18th centuries, it did not matter too much when you turned up somewhere; in general terms half-an-hour here or there was no problem.

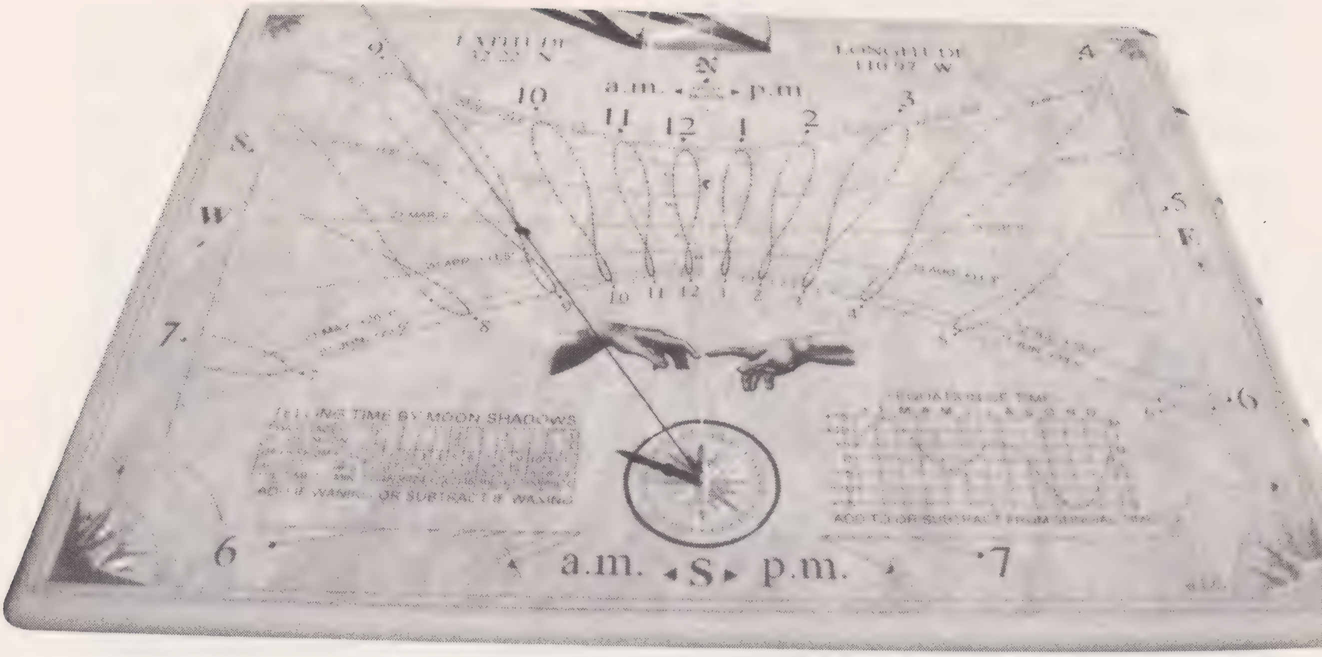


This sundial would make a charming wall feature but offers little in the way of accuracy.

Modern life leads us to demand more and more precision, and sundials are attempting to move with the times. As a rule, everyone knows if we are in summer time or not. The media takes care to warn us of the right day to put the hour back or forward, so the correction is a simple matter.

The location's longitude can be included, so that the person reading the sundial can make the correction themselves. It just has to be remembered that one degree is equivalent to four minutes of time (therefore 15 degrees is equivalent to one hour), as explained above.

In many cases, the equation of time is also included one way or another, with a table, a graphic or by adding the figure-of-eight of the analemma on one or on all of the hour lines; some sundials have it incorporated into the gnomon and, in this way, it is corrected always and at all times, though reading the time can be awkward if you don't know how to do it.



A sundial that gives information on the longitude and incorporates a graphic with the equation of time. The analemma is also shown on all the hour-lines.

SPOT THE MISTAKE

The horizontal sundial in the photo is incorrectly set up. Can you spot the mistake? In case you should fail to spot it, it is this: as we pointed out above, the north-south line should coincide with the hour-line 12:00, but the compass tells us that it is not in this case, so the sundial could be showing any time at all. It would appear that it takes a lot of patience to be a good sundial!

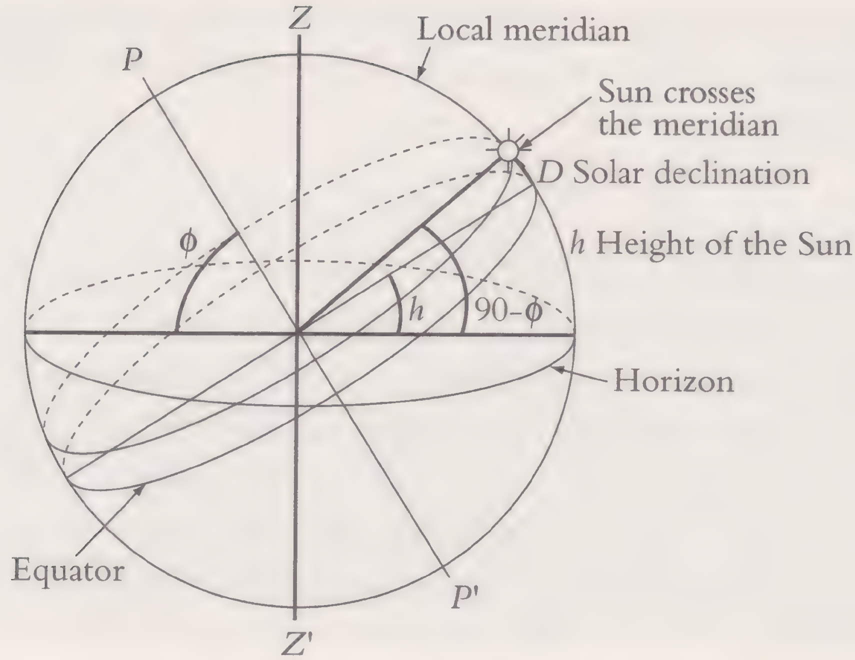
After all these changes, sundials really ought to be better appreciated by the public at large. There are of course ongoing problems. A sundial is not a flower pot and it has to be correctly positioned.

Longitude: a problem of time

A fleet of five English ships set sail from Gibraltar to begin their voyage home. Near the Isles of Scilly, south-west of the coast of Cornwall, Admiral Cloudesley Shovell calls his officers together to discuss the position of the fleet. It is 22 October 1701, a misty night with very poor visibility. They all agree that they should continue sailing north. Legend has it that a sailor on the flagship, HMS *Association*, informs the admiral that according to his own calculations the bearing is wrong, as their position has been poorly calculated. Discipline on board is very strict, and the sailor is immediately hanged for insubordination. A few hours later, *Association* hits the huge reefs off Scilly and sinks within minutes. Three other ships from the fleet hit the rocks and are likewise doomed, and only one manages to get away. More than two thousand men die in the disaster.

This true story is not unique; this sort of situation happened over and over again. It was a serious problem attempting to navigate far from the coast because mariners had no means of determining their position. For centuries, any experienced sailor had been capable of working out his latitude, both by the height of the North Star – if he was in the northern hemisphere – and by the passing of the Sun through the solar noon, that is, over the location's meridian. But the big problem lay in determining longitude. As has been explained earlier, the height of the North Star coincides with the location's latitude. It can be calculated by means of other stars, such as the Southern Cross or Orion's Belt, though it is somewhat more complicated. Likewise, it is possible to determine latitude simply by observing the Sun's passing over the meridian at 12 o' clock solar time. The latitude of the location is obtained by measuring the height h of the Sun over the horizon, and by knowing the Sun's declination on the day the observation is made.

With such a simple instrument as a quadrant or a cross-staff, or a more modern one such as a sextant or an octant, it is a simple task to measure the height of the Sun over the horizon at the time it crosses the north-south meridian – that is, at the instant that the Sun is highest above the horizon. This angle h, as can be seen in the next diagram is $h = 90° - \phi + D$, where the declination of the Sun is tabulated for each day in the astronomical almanacs. By working it out, we get: $\phi = 90° - h + D$.



The height of the Sun in its passing over the local meridian is equal to the colatitude 90–ϕ corrected for the solar declination corresponding to its own sign. The declination can be positive or negative depending on whether it is summer/spring or winter/autumn.

But it is not at all simple to find out the longitude. Columbus, in fact, attempted to reach the Indies in 1492 by following the parallel from his starting point in the Canary Isles. He travelled all the time with a constant latitude, and if he did not manage to reach Japan it was because he came across the American continent. The method he was following avoided the problem of knowing the longitude. In actual fact all the crew would have perished if they had not arrived in America because Columbus had made a mistake in calculating the size of the globe. He thought it was much smaller and consequently their supplies were running low when they spotted the continent. They were lucky…

But let's look at why it is so difficult to know the longitude. As was mentioned above, the terrestrial rotation clearly defines the rotation axis and the Equator. Parallel to it we can imagine all the circumferences that we want, but they will all be smaller: the Equator is unique. But the meridians, great circles which pass through the poles, are all equal in length. Agreement can be reached on a prime meridian and, in fact, that is what happened when the time came, but the problem is in determining from any point on the terrestrial surface the angular distance to that prime meridian, which is a line that is not a product of any astronomical concept, but rather of an agreement of a political nature. That, basically, was the origin of the great problem of longitude. When, some centuries ago, ships left harbour, they only had one very

rudimentary method of calculating their position. Distance travelled was calculated by first ascertaining the ship's speed. This was done by throwing a weighted log (it floated, but remained stationary on the surface) overboard and counting how many knots tied at regular intervals in a length of rope attached to the log were pulled through a sailor's hands during 30 seconds, timed with a sand clock. Knowing their speed, the crew estimated their position. But a vessel's speed varies with the wind, currents, the rope used..., in reality, it was impossible to calculate an exact position. Voyages took months and the lack of vitamin C (due to insufficient vegetables and fresh fruit) led to irreversible health problems and sailors perished from scurvy. All the major powers were anxious to solve this navigation problem which, for 300 years, occupied the minds of the greatest scientists of the day.

As explained above, 15° equals one hour or, put another way, one degree of longitude is the equivalent of four minutes of time. On the Equator, where the Earth's circumference is greatest, one degree of latitude equals some 111km, a figure that reduces as you change latitude. In other words, a single minute means a miscalculation of latitude of almost 28km. However, to the north or south of the Equator, a degree of latitude decreases in distance, which gives rise to a complete lack of precision.

After sailing for months it was impossible to calculate the position of a ship. Captains did not dare to stray from established routes which resulted in ship traffic always being in the same areas, which meant pirates knew they would find their victims within a more or less limited zone. For instance, in 1590 the Portuguese ship *Madre de Deus* was intercepted by an English squadron which obtained booty valued at half a million pounds, the equivalent of half the English Inland Revenue's total annual income at that time. This problem had to be solved.

In response to this preference on the part of ship's captains for following known routes, in the 18th century a curious project emerged. The aim was to anchor ships in the Atlantic every 600 miles and fire canons and flares from them that would be visible from 100 miles away, thereby guiding sailors. The idea was to create a maritime safety highway, but neither this nor other bizarre projects were successful.

Carlos V and Philip II of Spain, George II of England and Louis XIV of France put great efforts into searching for a solution. Commerce with the West Indies, military expeditions and the desire to discover other territories brought about a great increase in sailing and, consequently, an increase in the number of accidents and shipwrecks, with the loss of numerous lives and large cargoes. The underlying problem behind the ignorance of longitude was the great distortions which, at least

until the 17th century, existed on all known maps. This created serious problems when the era's nautical charts were being drawn up, and consequently was the origin of serious disputes over maritime zones that had been incorrectly demarcated. This explains why numerous islands in Oceania were 'discovered' on two or three different occasions. A mariner would discover an island that was not on the map and take possession of it in the name of his king. A few years later, another sailor would 'discover' the same island and place it on the marine chart in a different place to the first mariner and, from there on, numerous problems would arise, particularly between French and English sailors who believed, often in good faith, that they had been the first to find the island in dispute.

Basically, two solutions were found: an astronomical method and a mechanical one. The basis of the astronomical method was to study the movement of the heavenly bodies so as to be able to compare their positions. The mechanical solution was based on constructing a good, reliable mechanical clock that would enable the time to be known with accuracy. The fact is that the problem of determining longitude is, in reality, a problem of time: a difference in time is equivalent to a difference in longitude. The difficulty lay in being able to measure that difference in time.

Every sailor who possessed sufficient knowledge was able to determine the solar noon with accuracy, but the problem lay in knowing the local time. If he could know how the solar noon corresponded to that of his port of origin, the difference in times would give the difference in longitude – one degree corresponding to four minutes. The difficulty consisted of being able to maintain the information on the time at the port of origin.

The astronomical way

Let's suppose that an observer is situated at the centre of the Earth and has a reliable timepiece. First, he observes the passing of a star over the meridian at the port of origin at time t_1; the Earth carries on moving around and he later observes the same star passing over the meridian of a particular location at time t_2. The difference in time $t_2 - t_1$ will correspond to the difference in longitude. However, instead of being at the centre of the Earth, the observer is on the surface, he can only observe the passing of the star over his local meridian. He will know the time of the Sun's passing over the original meridian from his astronomical tables, but the difference in times will provide the answer to the problem.

Let's look at how the issue of comparing the local observation with that of the port of origin can be resolved. One solution is provided by eclipses. Let's suppose that the observer is in the middle of the Atlantic and he sees an eclipse of the Moon. If he knows that that eclipse occurs in London at time h_1 and that for him it is time h_2, the difference $h_2 - h_1$ will give him the difference in longitude between his location and London. But how accurate is the time that he is reading as h_2 on his sand clock? And besides, there are not eclipses every night, and mariners need to find out their longitude once a day, at least.

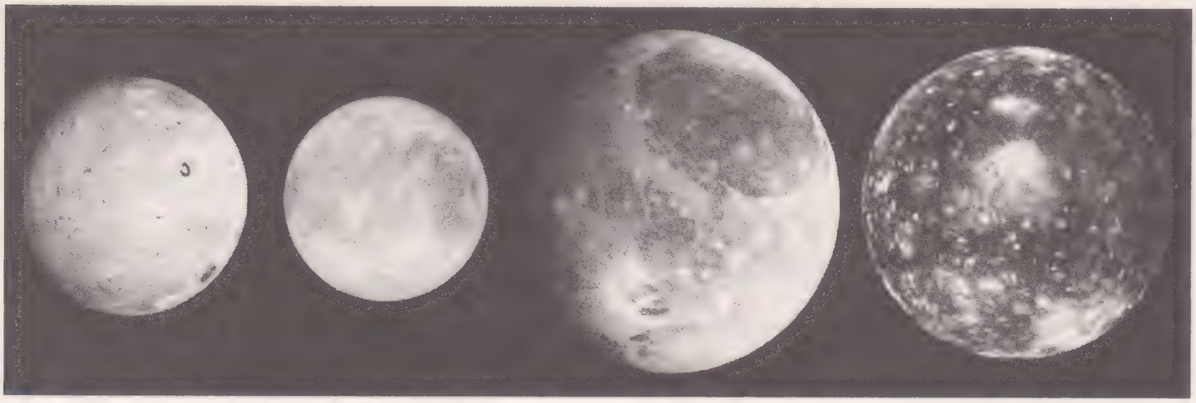
In 1514, Johannes Werner invented the method of the lunar distance, which was later perfected. We know that the Moon covers a distance which is approximately equal to its diameter, that is, half a degree, every hour. If we had a highly accurate celestial chart, then, when the Moon is 'touching' a known star, we could know at what exact time it does so at the origin meridian.

If the observer reads the moment at which the Moon touches that star from his ship, he can calculate the difference in times between that of the origin meridian and his own (which he knows from an imprecise sand clock). But there was a serious problem, in that at that time mariners did not know enough about the Moon's complex motion. Halfway through the 18th century, the lunar distance method began to be reliable. By then, mariners had more than 40,000 lunar and stellar observations carried out by John Flamsteed (1646–1719); Tycho Brahe had compiled an excellent atlas of the sky thanks to his talents for observation; Halley had made an accurate study of Earth-Moon influences, and Hadley had invented a quadrant which, with the aid of mirrors, enabled angular heights to be obtained over an artificial horizon in the event that the natural one could not be seen. This quadrant developed into a sextant with a small built-in telescope and a better range of precision.

In 1610, Galileo observed the satellites of Jupiter: Io, Europa, Ganymede and Callisto, which would later be known as Jupiter's Galilean moons. These satellites orbit around that giant planet, and on their orbits many fully predictable eclipses occur. Galileo proposed the result of his observations as a method of solving the longitude problem. However, this observation is difficult even from an observatory on land, so the method was not feasible. Even though Galileo went so far as to invent a helmet fitted with a telescope to make the observation easier, he admitted that even the observer's heartbeat could affect the viewing of the eclipse, making the results imprecise. After Galileo's death, and with the improvements made in telescopes, this method was used on land to be able to calculate longitude with more precision and,

consequently, to map countries with greater accuracy. Louis XIV even said that he "was losing more territory through his cartographers than through his enemies".

Another method considered at the time was the variation in the magnetic field, but it had to be rejected as the variation is not constant; it is dependent on location and time.



Jupiter's Galilean satellites: (from left to right) Io, Europa, Ganymede and Callisto (source: NASA).

ECLIPSES OF THE GALILEAN MOONS AND THE SPEED OF LIGHT

In the late 17th century, the renowned Italian astronomer Giovanni Cassini published his tables of Jupiter's Galilean satellites forecasting their eclipses. By taking into account that the Earth orbited around the Sun, a young Danish astronomer, Ole Rømer, found that the eclipses happened before the time expected when the Earth was nearer Jupiter and were delayed when it was further away. From that fact he deduced that these errors were due to the time that light takes to cover the diameter of the Earth's orbit. The speed of light could therefore be worked out. Rømer came up with a speed equivalent to 220,000km/s, somewhat lower that the true speed of 299,792km/s, but an incredible achievement nevertheless.

The solution was in the hands of a clockmaker

It was essential to create a good mechanical clock that could help sailors on their travels. But more than that, if a good clock could be found to mark out the time, mariners could carry it in their ship and keep the time of the port of origin (with a known longitude). If it kept good time, they would only have to take a reading of the Sun at its passing over the local meridian (at noon, 12:00), to read the time on

the clock (which is keeping the time at the port of origin), and then to subtract one from the other. This value could be added to the longitude of the port of origin and they would know the longitude of their location. Just by doing a simple subtraction! It was Huygens who was the first to build a pendulum clock, but it needed good atmospheric conditions for it to work, and therefore a pendulum clock could not be used on a ship. Huygens himself later invented an alternative to the pendulum: a spring-loaded wheel, but mariners still had to wait quite a number of years before the problem could be solved.

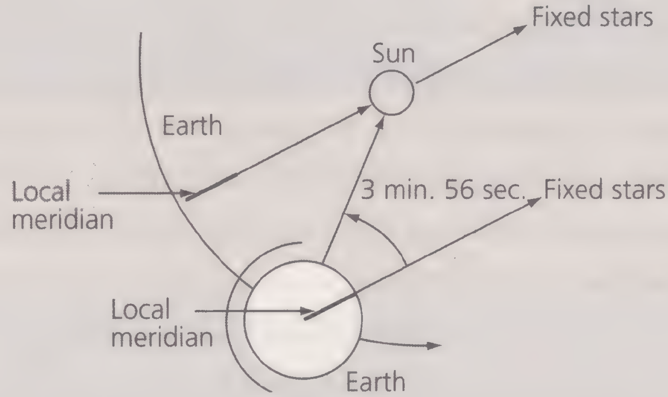
Although other countries also offered prizes (for example, Philip III, the king of Spain, offered a life pension for anyone who could solve the problem of longitude; in actual fact his advisers rejected Galileo's method of Jupiter's moons), the most attractive reward was offered by England under the reign of Queen Anne. Thus, in 1714 the British parliament passed the Longitude Act offering a prize of £20,000 to the author of a method of solving the problem with a margin of error of half a degree (ie, 50km at the Equator). As can be seen, that is not a high level of precision, but the prize was huge, which clearly shows the desperation of a country which was seeing its economy endangered if it could not provide its ships with a solution to this problem. To win the prize, the method proposed had to pass the test of accompanying a naval ship on its voyage from the England to the West Indies and back and obtain the required results. To oversee the process a Board of Longitude was set up, which included the Astronomer Royal, the head of the Royal Observatory at Greenwich, the director of the Royal Society, the minister in charge of the Navy, the leader of the House of Commons, a representative of the Army and several scientists. There were many who took part in the battle to win the Longitude Act's prize. We shall just highlight Jeremy Thacker, who created a timepiece which offered two developments that have survived through to our times: a vacuum chamber encased in glass and a system which allowed the device to be wound up without stopping the machinery.

The final solution arrived thanks to a carpenter, John Harrison, who built his first clock out of small pieces of brass when he was still not 20 years old, having taught himself. He added a table for the equation of time to the clock so that comparison with the solar time could be made. He designed a pendulum made up of two different alternate metals so as to counter the expansion and contraction of metal caused by changes in temperature. He also managed to control and regulate the energy transmitted to the mechanism by the spring. In addition, to eliminate the need for a pendulum, he developed a system of cogs that drew their energy from a coiled spring.

SOLAR TIME AND SIDEREAL TIME

By general agreement, our clocks work to a day of 24 hours, which is the time the Sun needs to pass two consecutive times over a meridian. We have already said that, in reality, this is the mean solar time, as the Earth follows the Law of Areas in its path around the Sun. Sometimes the Sun moves faster and sometimes slower, but as an annual mean, the Sun needs 24 hours to make one complete revolution over our horizon and pass over our local meridian.

But, if instead of taking the Sun as a reference point, we take the fixed stars, the period of rotation is somewhat less. Any fixed star takes 23 hours 56 minutes and 4 seconds to pass two consecutive times over the same local meridian, as the Earth, in its orbital motion, will have advanced a small angle with respect to the Sun: 3 minutes and 56 seconds.



The difference between solar time and sidereal time.

From his rudimentary observatory, John Harrison made sure his inventions were perfect by taking the movement of certain stars as his basis for time, ie, stars that appeared 3 minutes and 56 seconds before they had done the previous night. In this way he was able to make clocks that had an error of only one second a month, when other clocks of that time would vary by a minute per day. On account of this, Harrison managed to get a loan from the Longitude Board to build his first marine clock, called H1. It took him five years to finish it. Made from wood and weighing 34kg, it was inside a glass box with sides one metre long; it still works correctly. It was embarked on a ship to Lisbon and proved itself to be extremely useful by preventing the captain from making a serious error. In 1737, the Longitude Board met for the first time and gave unanimous backing to the H1. It was only Harrison himself who found fault with it and he managed to get more subsidies from the board to make the changes he saw fit. And that was how, in 1739, he built H2, and then H3 in 1751. It was necessary to wait

for clock H4 to see a clear decrease in size and weight. It was, in fact, after receiving a pocket watch – a present from one of his pupils – that Harrison began to work on a timepiece that was definitively different. Harrison's H4 was 133mm in diameter and weighed 1.3kg; it worked continuously for 30 hours and did not stop when being wound. In October 1761, Harrison embarked on a voyage to Jamaica and, on arrival at Port Royal, after a voyage of two months, it was found by astronomical means that the clock had lost only five seconds, which corresponded to an error in longitude of 1.25 minutes (about 2,000m), much more accurate than required by the Longitude Act. However, the Board decided that the experiments carried out had not been sufficient to determine longitude at sea. The problem was that three new members, mathematicians, had joined the Board, and they claimed that the longitude of Port Royal had not been established by using the method of Jupiter's moons, but the captain had been unaware that he was expected to do so, and neither would he have known how to do it. Clock H4 was again embarked in 1764, and this time the Board concluded that "the clock functioned with sufficient correction". But the Board offered Harrison only half the prize and on condition that he deliver two more copies of the clock and disclose all its secrets so that it could be mass produced.

Harrison took three years to make one copy, H5, and he was now 79 years old, so he was not sure if he would finish the second copy. Fortunately, King George III urged the Board to pay Harrison the rest of the prize money (Harrison always maintained that he had still not been completely paid). Clock H5 had a small error of only one-third of a second a day. It was a marvel! With his clocks almost completely devoid of friction and not requiring lubricants, perfectly balanced however much they were moved, able to run constantly despite changes in temperature, it was Harrison, and not the mathematicians who, after 40 years of innovation, would manage to get his hands on the Longitude Act prize. It is impossible not to admire the ingenuity of his work.

Harrison died in 1776, and many clockmakers were able to start up building marine chronometers by copying his inventions. By 1860, the Royal Navy had 800 chronometers for 200 ships. Within a short time the chronometer became a common and essential feature of a ship. The consequence of this was British dominance of the seas and the continued creation of the British Empire thanks to the precision and speed in calculating ships' positions. The marine chronometer was the instrument that guided navigators until the appearance of modern systems of navigation based on satellites in space.

Satellites guide us: the Global Positioning System

GPS (Global Positioning System) is a global system for navigation by satellite which provides data on the position of any object on the terrestrial surface with high precision. This system is made up of 32 satellites at an altitude of 20,200km on synchronised trajectories able to cover the whole of the Earth's surface. Additionally there is a set of land stations which send information to control and maintain the satellites, plus the receiving devices, which enable us to know our exact position. For automatic localisation, each receiver uses signals from at least three of the network's satellites. Thanks to these signals, the device can synchronise the GPS clock and calculate the time that the signals need to reach the equipment.In this way it can calculate the distance by using triangulation, and, finally, work out true coordinates to display exactly where we are.

Chapter 5

The Great Periods of Time

In the previous chapter we looked at the measurement of time, that is, what is understood as 'ordinary time', the concept we have of time during our daily activities. Astronomy has served humankind by giving structure to this concept but, from the astronomical point of view, we must consider time in upper-case letters. The time that corresponds to astronomical phenomena and cosmological developments is an 'XXXL-sized' time, and at this level mathematics are essential. As observations only go back some 300 years, the process followed by astronomers to study the 'great periods of time' has to make use of IT simulations, and for that reason mathematics is indispensable to form models which, when appropriately compared with observations, allow us to deduce what happened and will happen within a range of several billion years. These are the time scales of astronomical phenomena.

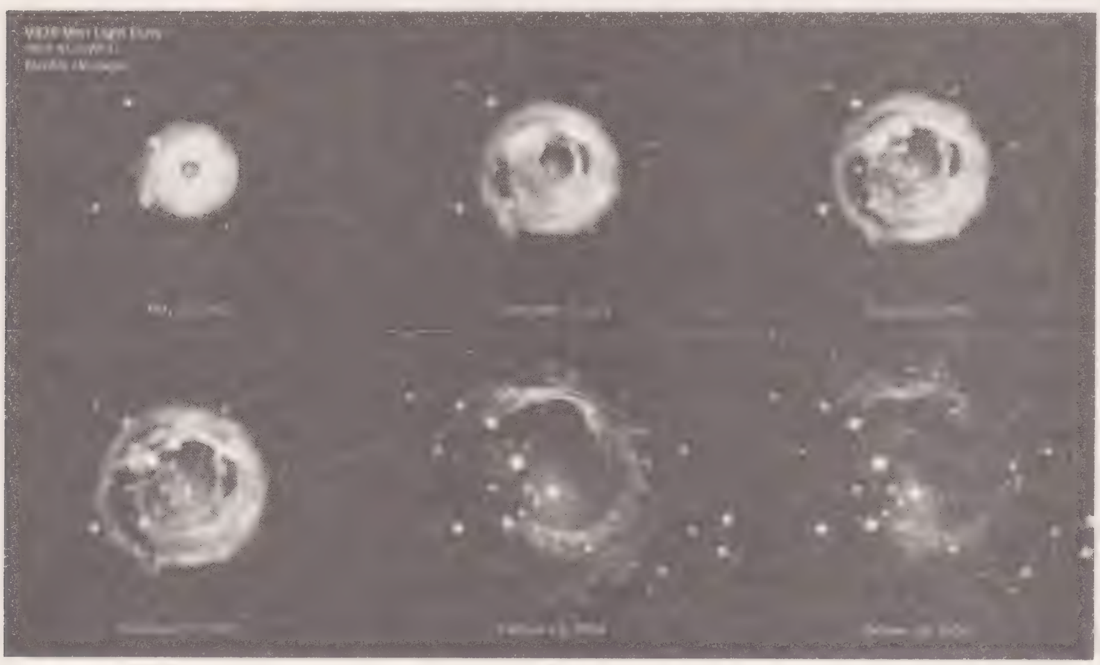
The stars, evolution and other characteristics

All stars are born within nebulae – clouds of gas and dust from the interstellar environment. Gravitational forces cause the clouds to contract and form stars in a process that lasts millions of years, because when they contract, clouds heat up and the contraction is halted until the cloud cools again through radiation. When, by means of interstellar winds, the new stars throw off the remains of the cloud that created them, the surrounding gas is illuminated and it shines brighter. Nebulae of many different colours can be seen in zones where there has been recent stellar formation.

In the origin of the Universe, nearly 14 billion years ago, the atoms that were formed were nearly all hydrogen and helium. The first stars were born from the contraction of clouds of gas containing solely these elements. The other heavier elements were generated throughout the evolution of those stars. After evolving and dying, they threw off part of their mass into space, which was now enriched with the new chemical elements that had been synthesised inside them, such as oxygen and carbon. Other heavier ones, such as lead and uranium, were formed during the explosion as a supernova.

The medium-sized stars, similar to the Sun, consume hydrogen in their nuclei and produce helium, which is later converted into carbon, nitrogen and oxygen. At the end of their existence, they expand and launch their atmosphere into space, forming a beautiful planetary nebula.

Nebula NGC 7635 shows a stellar wind bubble, an expanding zone around a hot star, which sends out a shock wave into interstellar space (source: NASA).



Images of a bright expanding zone that surrounds the V838 star in the constellation by Monoceros (the Unicorn). In 2002 it transformed into one of the brightest stars in the Milky Way for reasons that were unclear to astronomers.

CHILDREN OF THE STARS

We can think of a star as a giant nuclear power station in which nuclear reactions are taking place that convert hydrogen into heavier elements such as carbon, nitrogen and oxygen. All of these materials are necessary for life to exist. The atoms of the chemical elements making up the human body are formed in the bellies of the stars, which is why we can think of ourselves as offspring of the stars, and that is not just a line from a poem, but the truth!

Different generations of stars expel material which mixes with other clouds of gas in space, and eventually gives rise to other stars with their respective planetary systems. The concentration of heavy elements such as oxygen increases more and more in the interstellar environment, so much so that a way to determine the age of a star is to measure the amount of oxygen that it contains – the less it has, the longer ago it will have been formed.

The medium-sized stars, such as the Sun, consume hydrogen and produce helium in their cores, and later carbon and oxygen. At the end of their lives they expand and throw off their atmosphere into space. It is reckoned that within 4.5 billion years the Sun will become a giant red star and end its life as a beautiful planetary nebula. The Sun, the Earth and the other planets will be ejected into the interstellar environment and will serve as the foundations for the formation of new stars, with only a small white dwarf as a central remnant. The lifetime of a star depends on the amount of material that it possesses. The more massive it is, the shorter its life.

Stars smaller than the Sun do not eject material into space when their evolution comes to an end; they simply grow cold. But the most spectacular ends are those of the stars which are far more massive than ours. These stars explode as a supernova, material being flung out into space and dragging along the gas it meets, producing a spectacular cloud of hot gas. It is during this explosion that the heavier materials are formed, such as gold and uranium, while the nucleus compacts itself into a star of neutrons or into a black hole. Thus, in the different galaxies there are different phases of stellar formation, ranging from places where the birth of new stars has come to an end, to others where at this very moment thousands or millions of them are being formed.

How did the Solar System first appear?

The Solar System grew from an amorphous nebula which was rich in molecules of atomic and molecular hydrogen, helium, carbon dioxide, ammonia, water, refractory metallic dust grains and ice. It is thought that this pre-solar nebula was rotating in equilibrium in the interstellar environment and had a magnetic field. For some reason, perhaps caused by the shockwave of a nearby supernova, the pre-solar nebula contracted to a point where the gravitational force overcame the radiation pressure and it collapsed.

A depiction of the formation of the Solar System showing the forming star nucleus and the disk of surrounding material.

Due to the centrifugal force, it was easier for the cloud to collapse in the direction of the axis of rotation than in the perpendicular direction; therefore, the rotating cloud contracted into a disk perpendicular to the rotation axis in the same way that a ballet dancer's skirt forms the shape of a disk when she spins, ie, due to centrifugal force.

The denser particles of dust joined up to the plane of the forming system more quickly than the gas. The density tended to be greater in the centre of the cloud, thus forming the primitive sun, which conserved the magnetic field of the cloud that created it. The Sun condensed out of most of the available material. The gravitational energy of the original gas and dust were transformed into thermal energy when the cloud contracted and, consequently, the mass heated up. When its inner temperature reached several million degrees, nuclear reactions started up in its centre, and the newly born Sun became a star with its own light.

When the Sun began to radiate, it evaporated the non-refractory ice-dust that was nearby, but the same thing did not happen to the refractory metal dust, which survived the heating. Later, these specks of dust clustered together into particles and these, in turn, into ever-bigger rocks which, as they collided with each other went along trapping more and more material, thus creating the planets. The Sun's neighbouring planets are rich in heavy elements because they stem from refractory metallic dust which was not evaporated when the Sun was born. The most distant planets, on the other hand, have a chemical composition which is more like the original cloud, as they formed from refractory and ice dust. These outer planets are larger, and have rings and more satellites, which is because they had the chance to gather a larger amount of material when they were being formed.

A depiction of the protoplanetary disk with some forming planets (source: NASA).

A HOMEMADE MODEL OF THE FORMATION OF THE SOLAR SYSTEM

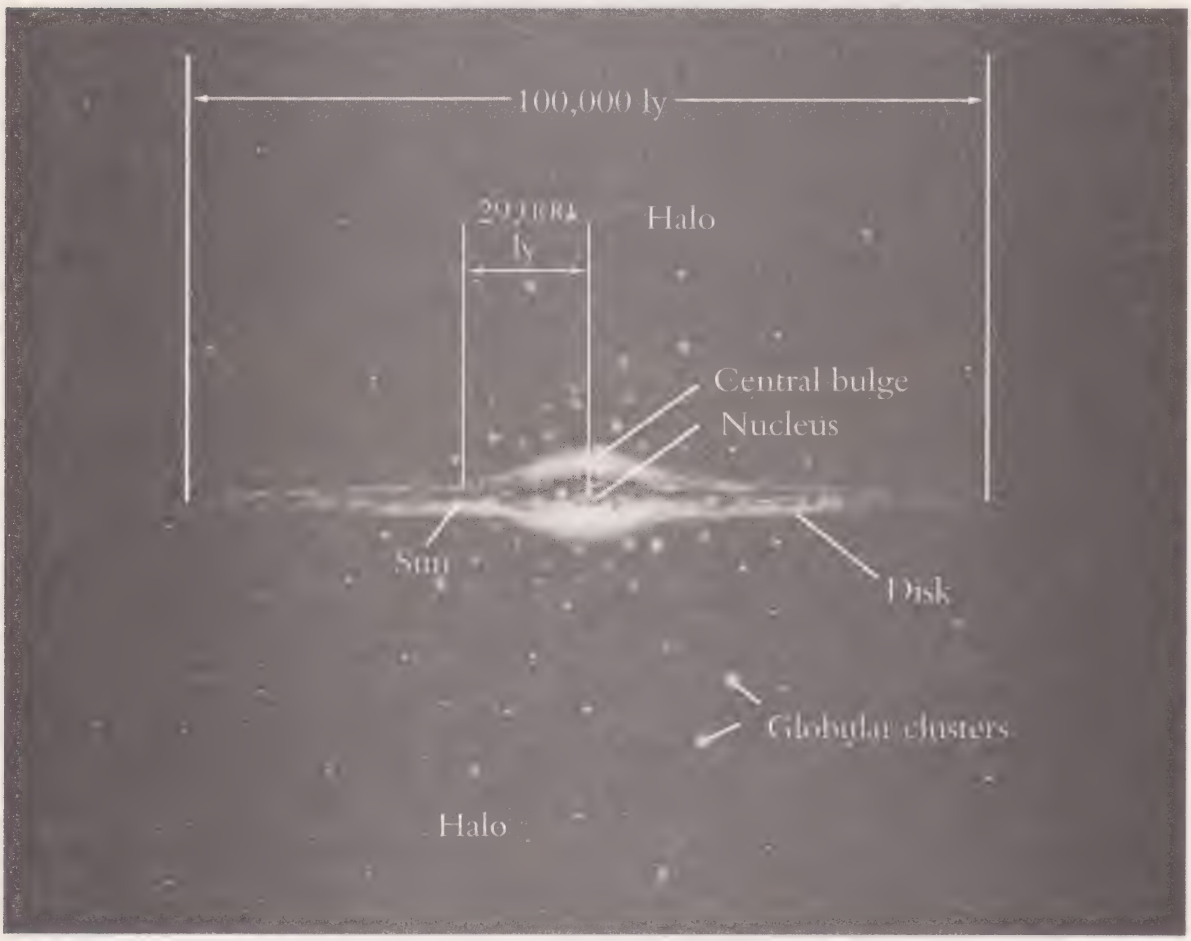
If we pour a small amount of oil into a bowl of water, the two liquids will not mix. Because they have different densities, the oil will float on the water and it has a tendency to form a single spherical body on the surface. If we move the water around with a spoon, creating a whirlpool, we will see how little drops of oil get dispersed, but they will later collide with each other forming ever-bigger drops of oil. Leaving aside obvious differences, this image is similar to what took place during the formation of our Solar System or any other planetary system.

Studying the stars

Stellar material is considered as a fluid that complies with hydrodynamic equations by adopting cylindrical symmetry, as that is what our galaxy displays. Using statistical and numerical methods, the superposition of two stellar populations is studied, which correspond to the recognised populations of type I and II. By using this superposition, the results concur with observations much more closely than if only one type of population is considered.

Historically, the difference between disk stars and those of a central bulge was discovered by the German astronomer Walter Baade (1893-1960). He defined the two classes of star as those of population I and those of population II, even before the stellar evolution process was known.

The classification criteria include the speed in space, localisation within the galaxy, age, chemical composition and differences in luminosity and colour.



The structure of the Milky Way showing the halo, the disk and the central bulge.

Stars of population I are those that Baade associated with the galaxy disk. They contain significant quantities of elements heavier than helium, which are the elements astronomers call 'metals'. These heavy elements were created in stars belonging to previous generations and were spread around the interstellar environment due to the explosion of supernovas. Our Sun is a star of population I. This type of star is common in the spiral arms of the Milky Way and in any other spiral galaxy.

The stars of population II belong to the first generations of stars created after the Big Bang and therefore most of them are not abundant in metals. It is unlikely that they will have planets orbiting round them. These stars are found in globular clusters and in the nucleus of the Milky Way. In the galaxy, the stars of population II of the stellar halo are of low metallicity, and are much older than those of population I. Knowledge of stellar evolution tells us that population II stars are of low mass, as the massive stars which were born together with the stars have already died.

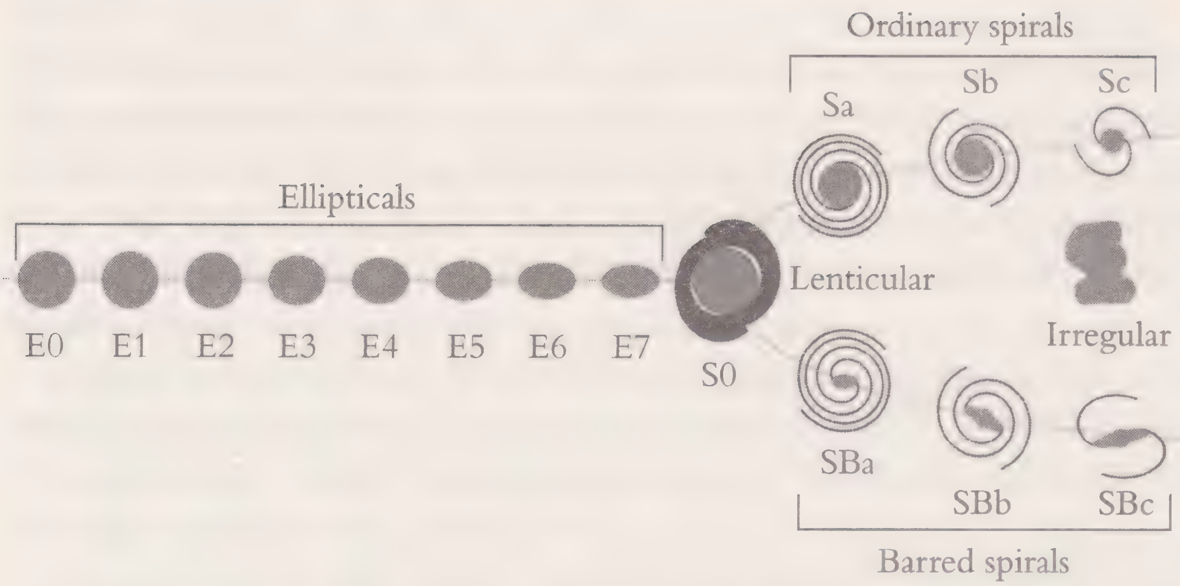
In 1925, the US astronomer Edwin Hubble classified the galaxies by following a scheme that is still used and which, in some way, helps to provide a simple explanation of their evolution. This classification system has two main categories: spiral and elliptic. But another two types are also differentiated: lenticular and irregular.

The spiral galaxies are basically divided into ordinary and barred according to their shape and the relative size of the bulge. They are characterised by the presence of a large amount of gas in the disk, which means that it has a high rate of star formation, particularly young ones of population I. They are normally found in zones with low galactic density.

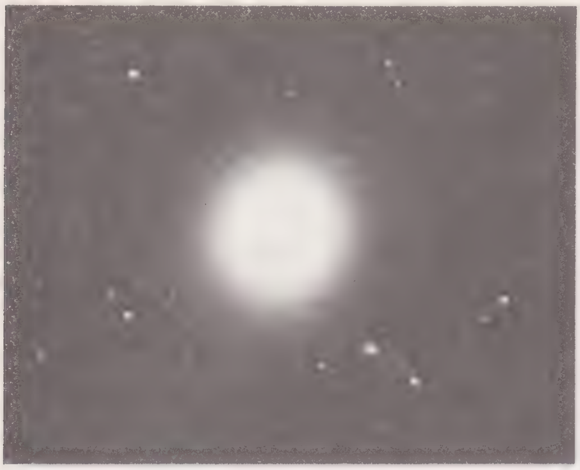
The elliptic galaxies are classified according to how rounded or elongated they are. They range from spherical (type E0) to very squashed ones (type E7). They display uniform brightness and are of the same appearance as the bulge without the disk of a spiral galaxy. The stars forming them are old, of population II, and contain practically no gas. They are usually found in zones of high galactic density, mainly in the centre of rich clusters.

Lenticular galaxies (type SO) which have a bulge and a disk but no spiral arms have now been added to the Hubble classification. They contain little or no gas, and are therefore old stars.

The irregular ones are small galaxies without a bulge and with a poorly defined shape. The Magellanic Clouds are the prototype of this kind of galaxy. Hubble proved that the galaxies are moving away from us at a speed proportional to their distance from us, in other words, Hubble proved that the Universe is expanding.



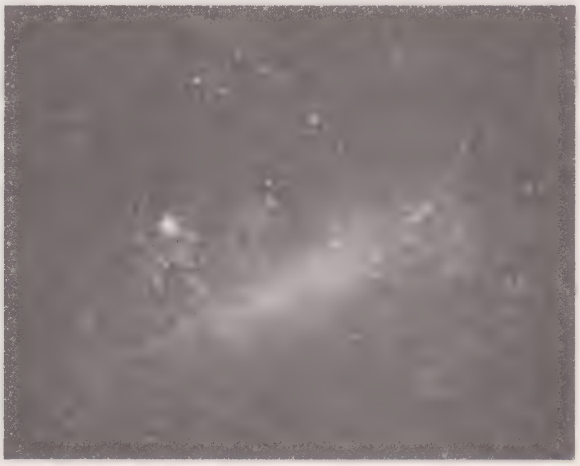
Above: Hubble's 1926 classification of the galaxies and their different subtypes (source: Ikonact).



Elliptic galaxy M87.

Spiral galaxy NGC 2997.

NGC 1365, a spiral galaxy situated at a distance of 56 million light years.



The Large Magellanic Cloud, an example of an irregular galaxy.

A GALAXY AT HOME

Spiral galaxies are the most spectacular ones. You can make your own representation of a spiral galaxy at home by using material which is very easy to get hold of. Take a round mould of the type used for making cakes (you will need to wash it aferwards!) and put in some water and a spoonful of very fine sand or grains of dirt (finely sieved soil will work). When the mixture is stirred with a spoon, the result is a lookalike of a spiral galaxy. It is a simple experiment that shows how material in rotation in space (and your home) forms spiral arms.

Stellar magnitudes and logarithms

Hipparchus of Nicaea, in the 2nd century BC, was the first to conceive of a method for classifying stars according to their apparent brightness. In his *Almagest*, Ptolemy published a classification system that followed the same criteria as Hipparchus' system, but which became more widely known. The brightest stars in the sky were denoted as being of first magnitude; those which displayed approximately half that brightness were known as second magnitude, and so on up to the sixth magnitude, which corresponded to those that could barely be seen with the naked eye (those which are only visible in exceptional conditions, when there is no moonlight and no light pollution). It was William Herschel (1738–1822) who noted that, on average, the light intensity of the first magnitude was a hundred times greater than that of the sixth; that is to say, to get the apparent brightness of a first magnitude star, it would be necessary to put together a hundred of the sixth magnitude.

In the 19th century, Norman Pogson (1829–1891) determined that the ratio between the luminosities of one magnitude and the next should remain constant, and he introduced a method that was more precise than the previous one and which is still used. He related the difference of apparent magnitudes of two heavenly bodies to the observations made on the terrestrial surface by supposing that the stars of the 6th magnitude were 100 times less bright than those of the first. Pogson defined the scale by establishing that the difference of five magnitudes corresponded to a factor 100 of brightness. So we have:

$$100^{1/5} = 2.512,$$

so the ratios between the magnitudes are the following: the first one is 2.512 brighter than the second; the first one is $(2.512)^2 = 6.31$ brighter than the third; the first one

is $(2.512)^3 = 15.85$ brighter than the fourth, and so one until the first magnitude is $(2.512)^5 = 100$ times brighter than the sixth. In other words, the sensation increases in arithmetical progression as the stimulus increases in geometrical progression. To sum up, for two stars of magnitudes m and m' the following ratio is verified for the respective luminosities B and B':

$$(2.512)^{m'-m} = \frac{B}{B'}.$$

Or, what comes to the same thing, if logarithms are used:

$$m' - m = 2.5 \cdot \log\left(\frac{B}{B'}\right),$$

where it suffices to consider that:

$$\log(2.512) = \log(100^{\frac{1}{5}}) = \frac{1}{5}\log(100) = \frac{2}{5} = \frac{1}{2.5}.$$

It is curious to note that our eyes respond to light logarithmically. That is, if a star shines 100 times more brightly than another one, it only seems to us to be five times brighter ($5 = 2.5 \log 100$).

The modern system is not limited to six magnitudes, and really bright objects have negative magnitudes. For example, Sirius, the brightest star in the northern hemisphere, has an apparent magnitude of -1.44 to -1.46. The modern scale includes the Moon and the Sun; the Moon has an apparent magnitude of -12.6 and the Sun, of -26.7. The Hubble telescope has found stars with magnitudes of $+30$. By considering that the brightness of a heavenly body is inversely proportional to the square of the distance separating them from us, we have:

$$m' - m = 5\log\left(\frac{D'}{D}\right),$$

where the distances of the two stars from us are expressed in parsecs.

Note that the apparent magnitude is not the same as the object's real brightness. An extremely bright star may appear really weak if it is very far away. So, to be able to compare the stars, the apparent magnitudes are not used, and instead the new concept of absolute magnitude is used. Absolute magnitude M is the magnitude with which we would see a star of visual magnitude m if it were situated exactly 10 parsecs from the Earth:

$$M - m = 5\log\left(\frac{10}{D}\right) = 5 - 5\log D.$$

In this way, the absolute magnitudes and the brightness of two or more heavenly bodies can be compared, and the distance makes no difference at all.

The war of the galaxies

Stars are born, develop, and die; galaxies evolve and interact. The Universe cannot be said to be unchanging. Let's look at some more things about the galaxies; in particular we shall study them a little more closely regarding their most dramatic interaction: the collisions of galaxies. In April 2008, NASA and ESA gave out a series of photographs taken by the Hubble Space Telescope which showed the ongoing 'war' of the galaxies in the Universe. They are truly impressive photos, though it must be pointed out that collisions between stars are very rare. They basically consist of a collision of huge masses of gas which cause an increase in the birthrate of new stars. Newborn stars which are big enough will evolve rapidly and explode as supernovas in just a few million years, with the heavier elements created in their interior being ejected and enriching the gas in the surroundings. In reality, these collisions are simply a new beginning.

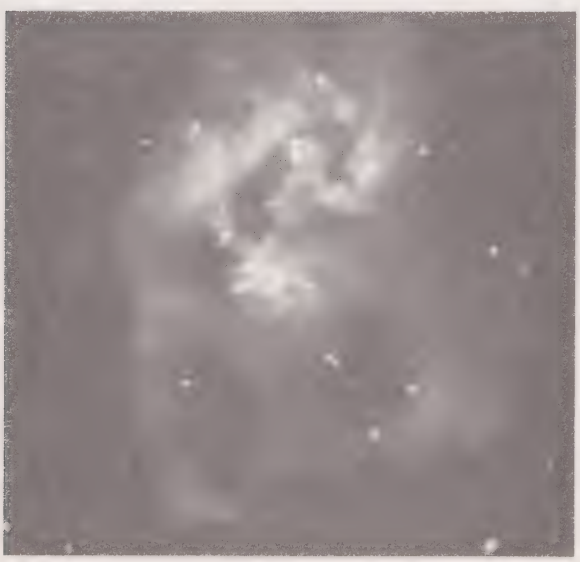
A photograph taken by the Hubble Space Telescope of the collision of two spiral galaxies: NGC 2207 (the bigger one) and IC 2163. The gravitational forces of the former have twisted the shape of the smaller one while ejecting stars and gases in long streamers that extend for up to 100,000 light years (source: NASA, ESA).

Hubble, again, has provided enough data for it to be confirmed that even the Milky Way conserves remains of old encounters of this type, and has enabled it to be seen that, at the present time, our galaxy is absorbing the dwarf elliptical galaxy of Sagittarius.

In April 2009, Hubble again offered us an image of a collision, but in this case it was at a much more advanced stage. NGC 2326 is a spectacular galaxy, with two long arms, situated 250 light years away in the Cancer constellation. The photograph below shows the final stage of the merging of two galaxies, with the two central nuclei merged into one. As they approach, great masses of gas pass from one galaxy to the centre of the other until they finally merge into a single structure. One single nucleus can be seen in this case, with two great tails formed by young stars, as the exchange of material has started off numerous processes to form stars which at present are in the first stages of their evolution. In this case, a black hole in the centre is aiding the merging process. The energy issued heats the disk and emits in different wavelengths.

Galaxy NGC 2326, a product of the merger of two galaxies. In this case the galactic nucleus is very active and is thought to be acting as a massive black hole which attracts material from its environs (source: Hubble, NASA-ESA).

The nearest example to us of a collision of two galaxies is the Antennae Galaxies system, which is only 45–65 million light years from us. The collision is thought to have started some 200 million years ago and was so violent that the gas and young stars were ejected in two insect antennae-like arcs from which the galaxy earned its name.

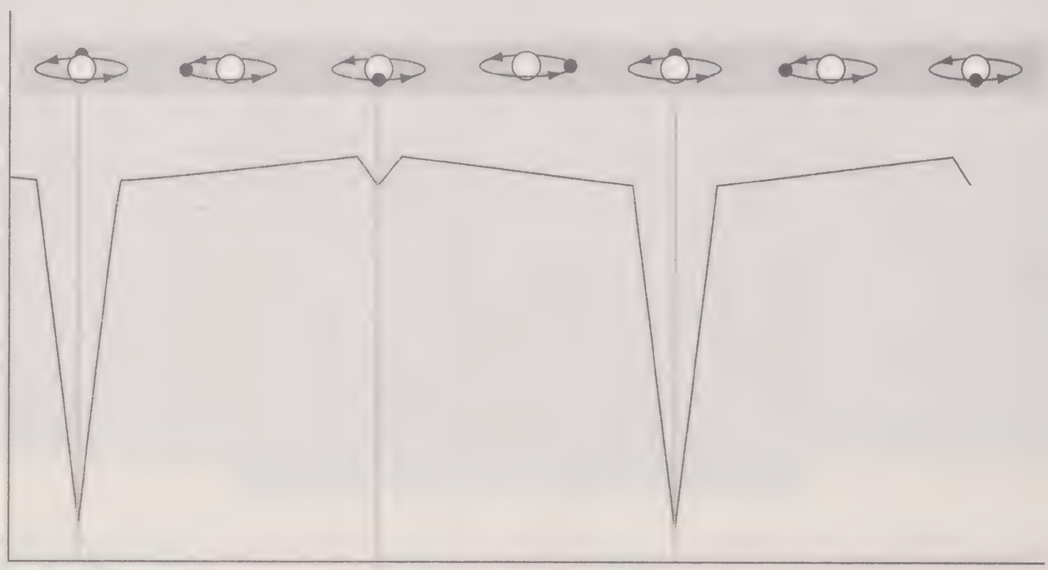


The Antennae Galaxies, the nearest collision to Earth that we can observe (source: Hubble NASA-ESA). The lower photograph is the image obtained by the Chandra X-ray Observatory revealing the existence of clouds of gas at a temperature of millions of degrees, neutron stars and black holes.

As the enriched gas formed during the collision cools down, new generations of stars and planets are being created. Recent research seems to indicate that stars with planetary systems are more likely to be formed in clouds enriched by heavy elements, and for that reason, in the future, there could be an unusually high number of planets

VARIABLE STARS AND THEIR LIGHT CURVES

Variable stars are those that display changes in magnitude. They can be like the supernovae that appear in the sky unexpectedly, but others undergo periodic changes and have been observed since ancient times. For example, Algol, the star β-Perseus, got its name from the Arabs, who had observed its spectacular changes in magnitude and so they christened it the 'Demon Star'. Approximately every two and a half days it changes its apparent magnitude from 2.2 to 3.3. Algol takes five hours to reach that minimum, and for 20 minutes stays like that, and then takes five hours again to reach its maximum. Algol is a typical example of an eclipsing binary. Its primary component is a blue star which is hot and brighter, while the secondary component is less bright and colder. Two types of eclipse can be seen depending on which star appears in front of the other. When the colder and less bright component moves in front of the main and hotter component, the brightness of the whole decreases; when the hotter one passes in front of the cold one there is another eclipse, but in this case the change in brightness of the whole is less notable. In all cases, the luminosity from both stars adds up, and the whole has a more or less constant brightness. Consequently, the brightness curve shows two minimums, as can be seen in the figure.



in the Antennae Galaxies. A great number of stars like the Sun and planetary systems similar to ours could grow old together for billions of years. If life should emerge in only a fraction of these planets, the Antennae Galaxies could be full of life in the future. The Antennae Galaxies offer a vision of the type of collisions that took place in

As has been said, one of the current fields of research in astronomy is the detection of extrasolar planets. There are several methods of detecting them; one is to study the changes in a star's brightness. A planet may temporarily 'darken' (eclipse) the star it is orbiting, causing a small decrease in brightness during the transit of 1% in the case of a giant planet such as Jupiter, and 0.01% if it is a planet the size of the Earth. One disadvantage of this method is that the planet's orbit has to be suitably aligned for it to produce an eclipse visible from Earth. The odds of this happening are approximately 0.5% for a planet situated at one astronomical unit (AU) from its star. In other words, if all the stars have a planet at a distance of 1 AU, it is necessary to follow 200 of them to be able to see an eclipse. If 10% of the stars have a planet at 1 AU, in order to detect 10, some 20,000 stars would have to be followed photometrically.

The Hubble Space Telescope, for instance, observed that HD 209458b produces an eclipse every four days. Spectroscopic observations taken during the eclipse have provided clues as to the chemical composition of its atmosphere. It should also be pointed out that planet OGLE-TR-56b, which was detected by this method, has eclipses every 30 hours, which implies, if the interpretation of the observations is correct, that the orbit is very small (five times the radius of the Sun). All of these are giant planets and not able to support forms of life similar to Earth's.

But not all the variable stars are of this type of eclipsing star. There are some that are intrinsically variable. For instance, δ-Cephei is one – its variations in light come from the expansion and compression of its internal structure. We can observe these changes and classify them according to their characteristics. They are normally known by the name of the first star of this type to be studied. Pulsating variable stars are put into several categories: δ-Scuti, RR-Lyrae, δ–Cephei and W-Virginis (here they are in order according to their periods, which are 0.10, 0.57, 5.34 and 17 days respectively). Cepheid stars are of great interest because they display a ratio between their period and their brightness, which provides the chance for their distances to be calculated.

the hectic early Universe and which probably gave rise to the formation of many of the stars that exist at the present stage of the Universe. They can also give us a preview of the future of our galaxy, the Milky Way, if it is involved in a possible collision with Andromeda.

Collision of the Milky Way and Andromeda

It is thought that within 3 billion years there will be a collision between the two major galaxies in the Local Group – the Milky Way and Andromeda – and both will merge into a larger and possibly elliptical galaxy. The two galaxies are approaching each other, in the case of the Sun at a current speed of 300 km/s. As the tangential speed of Andromeda is unknown, it is not known when the collision will take place or if the galaxies will just come closer together. But the fact is that there is very likely to be a collision sooner or later.

It was in 1959 that the idea of this possible collision was first mooted, but it has only been recently, and thanks to IT simulations using mathematical models, that it has been possible to research what would happen. Andromeda and the Milky

THE LOCAL GROUP

As mentioned above, galaxies group together in clusters. The Milky Way belongs to the cluster known as the Local Group, which includes some 30 galaxies. The two major ones are the Milky Way and Andromeda, which have many others as satellite galaxies. In fact, these two galaxies and the other free galaxies revolve around a central point. The Milky Way's disk is about 100,000 light years wide. Two of its satellite galaxies can be seen from the southern hemisphere without the aid of a telescope. They are the Large and the Small Magellanic Clouds, named after the 16th-century Portuguese explorer Fernando de Magellan, who brought back news of it to the northern hemisphere. They look like a pair of bright hazy marks in the sky, but modern photographic techniques show that they are made up of billions of stars. The Large Magellanic Cloud is 170,000 light years away, while the small one is at a distance of 190,000 light years. They are irregular-shaped galaxies situated not far from our own, which is the reason that they are fragmented and distorted due to the interactions of the gravitational tidal forces.

On the outskirts of the group lie a handful of isolated smaller galaxies. The largest of them is the galaxy named Triangulum, also known as M33. The Local Group is one of the many similar groups contained in the Virgo Supercluster.

Way seen from afar would look like galaxies NGC 2207 and IC 2163, and in time would begin to get further away and would be reminiscent of the Antennae Galaxies, with some differences. The two galaxies would separate from each other until their gravitational attraction put a brake on the process and forced them to approach each other again, finally clashing; the result would be a burst of stellar formation and the appearance of massive black holes in the centre of both galaxies; the galaxies would end up by merging into one, creating, eventually, what would possibly be an elliptical galaxy that has already been christened Milkomeda. The models even provide information on how the new galaxy would look. As all this activity is predicted to occur 3 billion years from now, it is thought that by that time there will be very little gas in the Milky Way and Andromeda, with the result that the birthrate of new stars will be lower than in other cases. Milkomeda would be a giant elliptical galaxy, though its centre would have a much lower density than normal. This object, together with its possible satellite galaxies would be the only things left of the Local Group. Following the collision, the most probable scenario is that our Sun would end up in the galactic halo of the Milkomeda, but if it occurs within 3 billion years, as expected, the Sun will be in the main sequence stage, although, according to the models of solar evolution, the Earth will be uninhabitable by then due to the increase of the Sun's luminosity.

The structure of the Universe

According to the latest observation data, the Universe looks like the mountain of foam you would create by pouring some washing-up liquid and water in a bowl and stirring it up with your hand. We shall leave aside the nebulae, planets, stars and other minor objects in the Universe and simply look at a Universe made up of galaxies as minor elements. We shall look at the distribution of galaxies in space forming groups called clusters of galaxies which, in turn, cluster together again. But let's see what the observation data are.

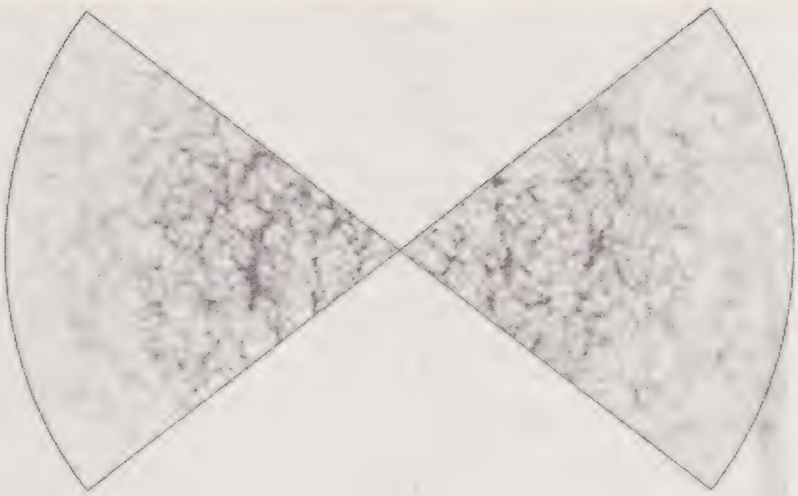
It is not easy to observe the layout of the galaxies, that is, the structure of the Universe on a grand scale, as we have no space vehicles enabling us to travel away from the Solar System. As we are anchored to the Earth and its small environs, we only have a two-dimensional vision of the Universe covering only what we see above in the celestial sphere. We need a 3D view and the only way to get it is by measuring the distance that separates us from each of the galaxies. This way we will get an idea of the galaxy in three dimensions.

Millions of galaxies compiled by using images from the UK's Schmidt telescope.

But astronomy's great problem, as said above, is being able to determine cosmological distances. The methods explained are no use for calculating them, and the most effective way is to measure their red shift. All galaxies have been moving away from each other since the Big Bang. The speed at which a galaxy recedes, or the speed at which it moves away from us, depends on the distance at which it is situated, in accordance with Hubble's Law. According to this astrophysicist there is a linear relationship between speed and distance, and the constant of proportionality is what is known as the Hubble constant, ie 71 km/s per megaparsec (a million parsecs), which was not at all easy to determine. By measuring a galaxy's recession speed, we can find out how far away it is by using Hubble's Law.

Discovering the recession speed of a galaxy is relatively simple – we just have to study its spectrum. Just as star light decomposes into a spectrum, it is possible to do the same with the light that reaches us from a galaxy. In this spectrum there are always some dark absorption lines that show the presence of chemical elements. The position of these lines on the spectrum depends on the recession speed of the galaxy. The greater it is, the nearer the red zone of the spectrum those dark lines will be; that is what is known as red shift.

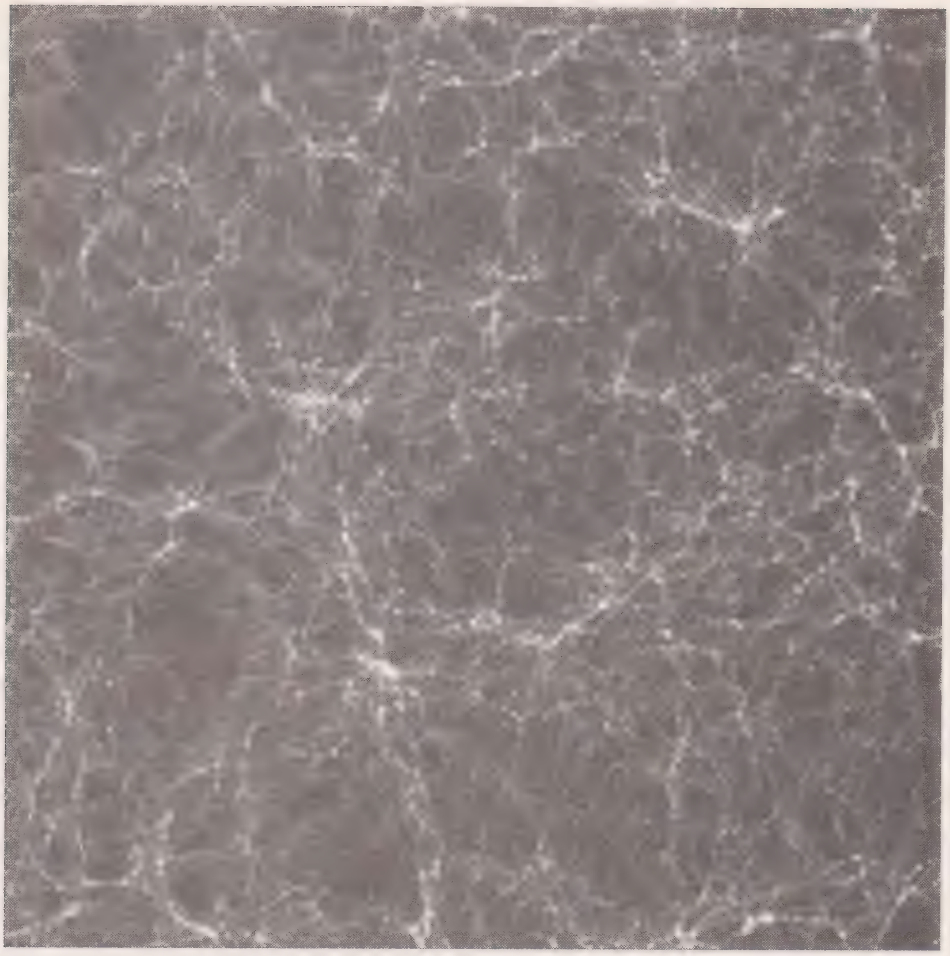
At the present time, there are several teams working to take the spectra of several million galaxies. The results obtained so far are still only somewhat partial. According to these results, the galaxies are grouped in filaments, leaving empty bubbles between them, in a structure similar to foam. The galaxies would be situated in the area of the soapy film, and the empty space would be the bubbles.



Two portions of the Universe studied on a large scale. Our position corresponds to the central vertex since the image reflects our point of view.

To make up a model to simulate the structure of the Universe it is necessary to start by considering its composition, and to bring into play mathematical equations to simulate the forces that are in operation. But determining that composition is one of the great obstacles that have to be taken into account. Apparently, there is a large amount of cold, dark matter, a matter that neither emits light nor blocks it, and as a result is truly invisible; however, its gravitational influence can indeed be measured. Also to be considered is the dark energy that seems to fill all space, and whose anti-gravitational effect is accelerating the expansion of the Universe. Dark energy and dark matter make up the majority of the composition of the Universe, and ordinary matter is only 4% of the total. With these presumptions fed into the Big Bang model, the simulations it produces have been shown to correspond to what we can actually observe. It seems incredible, but the virtual Universe can simulate billions of years of evolution, and the images they provide enable us to get a better understanding of what we are actually observing in the cosmos.

As the galaxies are not distributed randomly, the probability of finding one in a certain zone is determined by the average density and a correlation function of two points which, to a certain extent, describes the degree of concentration, as the concentration of galaxies is not independent of the zone in which it is situated. The type of function used in the statistical model is modified from time to time so that the simulations correspond to data obtained by observation. At the present time, this type of simulation is dealing with very advanced areas of research and different models are being used.



A detail from a simulation by the Virgo Consortium of the distribution of dark matter. More than 10,000 dots were used to created it.

By studying the redshift of the galaxy spectra, in 1986 sufficient variations were discovered to reveal the existence of a concentration of mass equivalent to thousands of galaxies situated at a distance of around 250 million light years in the direction of Hydra and Centaurus. This gravitational anomaly is known as the Great Attractor. This zone has large, old galaxies and many of them are colliding with nearby ones and emitting large amounts of radio waves.

In 1989 a galaxy cluster, the Great Wall, was discovered at a distance of more than 500 million light years. It is 200 million light years wide and only 15 million light years deep. In 2004, an empty superspace known as the WMAP Cold Spot was discovered almost a billion light years away in the Eridanus constellation. There are many other examples that confirm the appearance of bubbles in the Universe.

All these observations have to be regarded with caution. It has to be borne in mind that there may be a large number of errors and that things are not necessarily

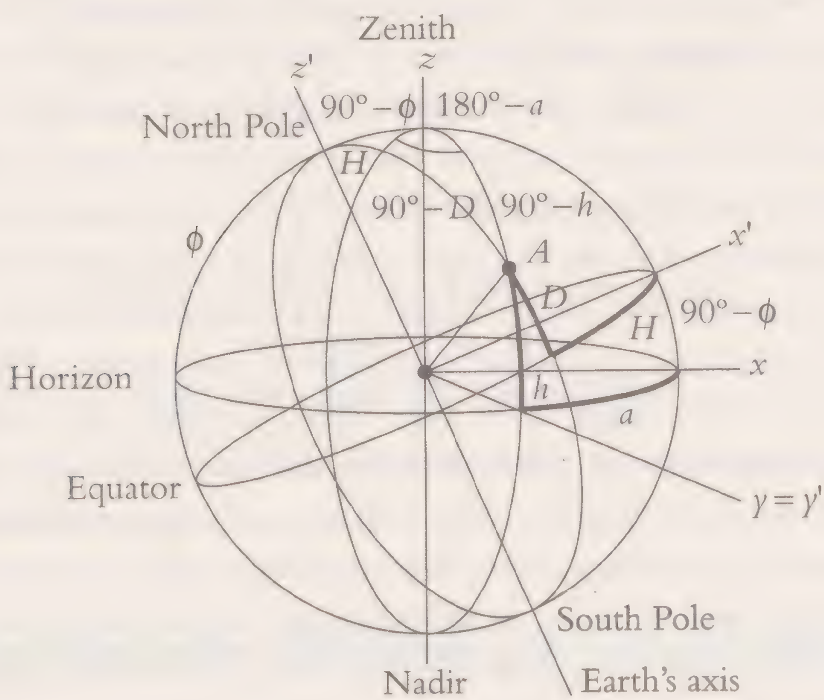
what they seem. Gravitational lenses can cause images that we receive to show positions that are different from actual positions. At the present time, images of great chunks of the sky are being compiled so that more can be learnt about the evolution of the Universe. The fact is that very large samples are needed in order to get good results, as otherwise a false impression can be gained. As pointed out above, there are several teams working on this subject and an increasing amount of information is expected to be gathered in the future, all of which will enable improvements to be made on current models.

Appendix

Who Wants to Learn More and Do Some Calculations?

Chapter 1: The conversion of coordinates and the triangle pole-zenith-star

The conversion between azimuthal and equatorial coordinates involves concepts of spherical trigonometry, and nowadays is done using conversion matrices.



Conversion of equatorial/hour coordinates to horizontal coordinates.

In the diagram above, the position of the celestial body A is determined by a vector whose three components correspond to the projection of the position of the body onto the plane of the horizon (plane xy) and the zenith-nadir axis (axis z); so, it can be expressed in reference to the three axis of coordinates x, y, z. Therefore, in horizontal coordinates, the position of A can be written as the vector $(r \cdot \cos(h) \cdot \cos(a)$, $r \cdot \cos(h) \cdot \sin(a)$, $r \cdot \sin(h))$. Analogously, we can go on to determine the position of the body in respect to the Equator (plane $x'y'$) and the axis of the Earth's rotation

(axis z'), that is, in respect to the axes x', y', z' in equatorial coordinates, and we have $(r \cdot \cos(D) \cdot \cos(H), r \cdot \cos(D) \cdot \sin(H), r \cdot \sin(D))$. As can be seen in the diagram, we can go from reference x, y, z to the reference x', y', z', simply by rotating with respect to axis y, which coincides with y' an angle of $90° - \phi$, this is the complementary of the latitude of the location, the co-latitude. Therefore, the axis x becomes the x' and, analogously, z becomes z'. The conversion matrix with respect to the second axis (axis $y = y'$) for an angle of $90° - \phi$ is:

$$\begin{pmatrix} \cos(90° - \phi) & 0 & \sin(90° - \phi) \\ 0 & 1 & 0 \\ -\sin(90° - \phi) & 0 & \cos(90° - \phi) \end{pmatrix}$$

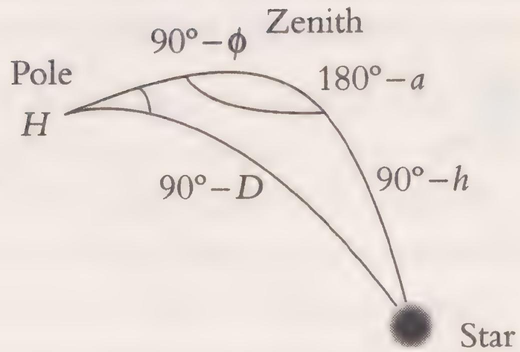
We thus have:

$$\begin{pmatrix} r\cos D \cos H \\ r\cos D \sin H \\ r \sin D \end{pmatrix} = \begin{pmatrix} \sin\phi & 0 & \cos\phi \\ 0 & 1 & 0 \\ -\cos\phi & 0 & \sin\phi \end{pmatrix} = \begin{pmatrix} r\cos h \cos a \\ r\cos h \sin a \\ r \sin h \end{pmatrix}$$

Therefore, the conversion formulae are:

$$\cos D \cdot \cos H = \sin\phi \cdot \cos h \cdot \cos a + \cos\phi \cdot \sin h$$
$$\cos D \cdot \sin H = \cos h \cdot \sin a$$
$$\sin D = -\cos\phi \cdot \cos h \cdot \cos a + \sin\phi \cdot \sin h.$$

The same relationships are obtained by using the conversion matrix as by applying Bessel's spherical trigonometry to what is known as the 'pole-zenith-star triangle', which is shown in the image on the following page. Astronomers have used this triangle for many years to make calculations on the positions of heavenly bodies. As previously there were no computers or calculators to do this work, logarithms were applied and use was made of the tables that directly gave the logarithmic values of the trigonometrical lines of the angles expressed in degrees, minutes and seconds. This spherical triangle was and is still used very much in positional astronomy, as it sums up all the information in the figure on the previous page. It should be borne in mind that the sides of the triangle are arcs of great circles on the celestial sphere and are therefore measured in degrees, though there is a tradition of giving the hour angle and right ascension in hours, minutes and seconds. The conversion to degrees, minutes and seconds is done easily by simply taking into account that 360° equals 24 hours, which means 15° is the same as one hour.



Pole-zenith-star triangle

Chapter 2: Aristarchus's calculation on the Earth–Moon–Sun system

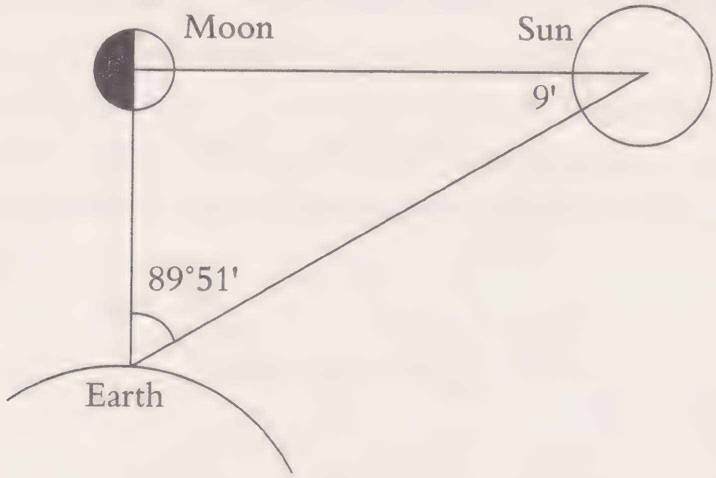
Aristarchus (310-230 BC) deduced the relationships between the distances and the radii of the Earth-Moon-Sun system. He calculated the ratios between the radius of the Sun, the radius of the Moon, the distance from the Earth to the Sun and the distance from the Earth to the Moon, and he determined it all in relationship to the radius of the Earth. Unfortunately, Aristarchus did not manage to establish the radius of our own planet and could not calculate the absolute values of all the radii and distances. It was Eratosthenes, a few years afterwards, who determined the radius of the Earth. By using modern notation (and modern-day values), let's look at the mathematical process designed by Aristarchus and we suggest that the reader repeat the experience of carrying out the observations and redoing the calculations to see what results they can obtain.

Relationship between the Earth–Moon and Earth–Sun distances

Aristarchus determined that the angle under which the Sun-Moon distance is observed from Earth at the time of a quarter moon was 87°. We nowadays know that he made an error, possibly due to the fact that he found it very difficult to determine the exact time of the quarter moon. That angle is, in fact, 89° 51', but the calculation process Aristarchus used was perfectly valid. If we use TS to denote the Earth-Sun distance and TL for the Earth-Moon distance with $\sin(9') = TL/TS$, we have:

$$TS = \frac{TL}{\sin(9')} = 400\ TL.$$

Aristarchus, in actual fact, deduced that $TS = 19\ TL$.



The relative position of the quarter moon.

The relationship between the lunar and the solar radii

The relationship between the radius of the Moon and that of the Sun must be similar to the formula obtained above, because from the Earth both diameters are observed as equal to 0.5°. Therefore, both radii verify:

$$R_S = 400\ R_L.$$

The relationship between the distance from the Earth to the Moon and the lunar radius or between the distance from the Earth to the Sun and the solar radius

As the Moon's observed diameter is 0.5°, with 720 times this diameter it is possible to cover the path, presumed to be circular, of the Moon around the Earth. The length of this path is 2π times the Earth-Moon distance, in other words $2\,R_L\,720 = 2\pi\ TL$. Simplifying,

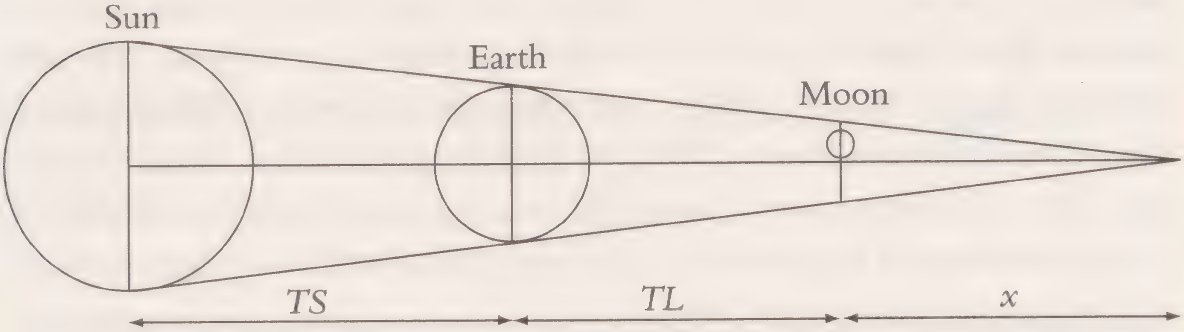
$$TL = \frac{720\ R_L}{\pi},$$

and by using a similar reasoning, supposing that the Earth revolves around the Sun describing a circumference of radius TS, we have:

$$TS = \frac{720\ R_S}{\pi}.$$

The relationship between the distances to the Earth, the lunar radius, the solar radius and the terrestrial radius

During an eclipse of the Moon, Aristarchus observed that the time required for the satellite to cross the cone of the terrestrial shadow was twice that required for the surface of the Moon to be totally covered by the shadow. He therefore deduced that the diameter of the Earth's shadow cone was double the diameter of the Moon, that is, that the ratio between both diameters or radii was 2:1. In reality, we know that this value is 2.6:1. Without doubt, during an eclipse of the Moon, by using a chronometer, it is possible to calculate the relationship between the following times: 'The first and last contact of the edge of the Moon with the terrestrial cone shadow.' In other words, to measure the diameter of the Earth's cone shadow, and 'the time required to cover the lunar surface', and a value close to 2.6:1 can be obtained without much difficulty.



Shadow cone and relative positions of the Earth–Moon–Sun system.

By following the diagram, the following proportions can be established, where x is a variable auxiliary whose only purpose is to facilitate the expressions:

$$\frac{x}{2.6\,R_L} = \frac{x+TL}{R_T} = \frac{x+TL+TS}{R_S}.$$

By introducing the relationships $TS = 400\ TL$ and $R_S = 400\ R_L$ into this system of equations, it is possible to eliminate the auxiliary variable x and, by simplifying, we obtain:

$$R_L = \frac{401}{1{,}440} R_T,$$

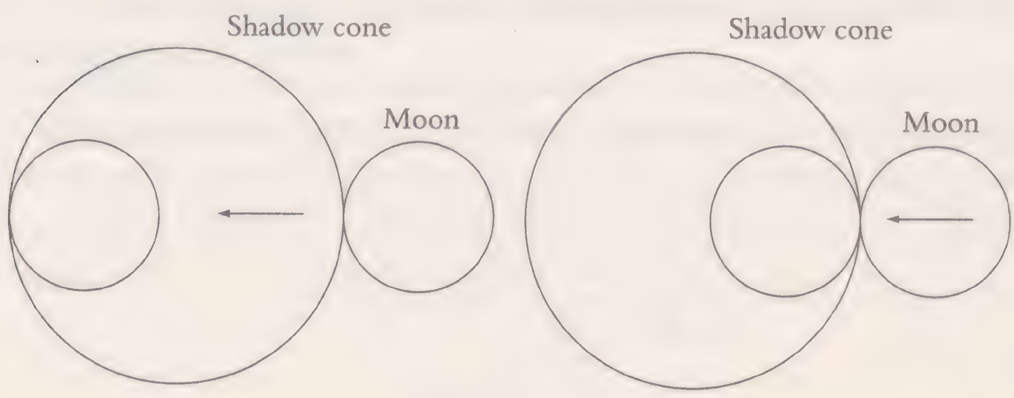
a formula which allows all the previously mentioned dimensions to be expressed relative to the radius of the Earth. Thus:

$$R_S = \frac{2,005}{18} R_T;$$

$$TS = \frac{80,200}{\pi} R_T$$

$$TL = \frac{401}{2\pi} R_T.$$

The only step required is to insert in the radius of our planet in order to find all the distances and radii of the Earth-Moon-Sun system. Aristarchus never managed to discover that value, and therefore he had obtained a set of ratios but was never able to calculate those distances and radii in explicit form. Today, the equatorial radius of the Earth is perfectly well known: it is 6,645 km, a value which, inserted in the expressions above, enables us to get the following results: $R_L = 1{,}850$km (real value, 1,738km); distance $TL = 424{,}000$km (real value, 384,000km), $R_S = 740{,}000$km (real value, 696,000km) and distance $TS = 169{,}600{,}000$km (real value, 149,680,000km). The most noteworthy aspect of these calculations is not the results themselves but in understanding the ingenious way that Aristarchus obtained such spectacular results with so few means at his disposal.



The first and last contact between the edge of the Moon and the Earth's shadow cone enables the diameter of the cone to be calculated (above left). The time required to cover the lunar surface enables the diameter of the Moon to be measured (above right).

Chapter 2: Determination of the mass of an extrasolar system's central star

If we take the movement of the exoplanets around a central star in a circular orbit of radius a, it can be formulated by equalising the forces that are acting:

$$\frac{m v^2}{a} = \frac{G M_E m}{a^2}.$$

By simplifying, the velocity v verifies:

$$v^2 = \frac{G M_E}{a}.$$

The period P for a circular movement is:

$$P = \frac{2\pi a}{v},$$

where by substituting the value of the previous velocity v , it is deduced that:

$$P^2 = \frac{4\pi^2 a^3}{G M_E}.$$

And for every exoplanet clearance is made of the constant that appears in Kepler's third law:

$$\frac{a^3}{P^2} = \frac{G M_E}{4\pi^2}.$$

By writing the above-mentioned relationship for the Earth, which revolves around the Sun with period $P=1$ year and radius of the orbit, presumed to be circular, $a=1$ AU, the following equation is deduced:

$$1 = \frac{G M_S}{4\pi^2}.$$

If the two latter equalities are divided, and by using the mass of the Sun as a unit, $M_S = 1$, we get:

$$\frac{a^3}{P^2} = M_E,$$

where we know that a is the radius of the orbit (in AU); P, the revolution period (in years), and we can determine the mass of the central star, M_E (in units of solar masses).

The mass of the central star M_E of an exoplanet system (in solar masses) can be obtained with the expression:

$$M_E = 0.0395 \cdot 10^{-18} \frac{a^3}{P^2},$$

where a is the radius of the orbit of the exoplanet (in km) and P is the revolution period (in days). As a particular example, the mass of the stars Ups And and Gl 581 can be calculated in solar masses and is given in the values that appear in the table on page 60.

Chapter 3: A simplified calculation of the Earth–Sun distance by using the 1769 transit of Venus

Though there is a risk of losing a little in terms of precision, the mathematical expansion has been simplified for those who are not specialists in celestial mechanics so as to provide an easily understood version based on the method proposed by Halley and Delile. Only two hypotheses have been adopted: to suppose that the orbits of Venus and of the Earth are circumferences centred on the Sun, and that the positions of Venus, the centre of the Sun and the terrestrial observers are coplanar. We shall therefore start off from those hypotheses and use the data obtained in the transit of 3 June 1769 for observers very distant from each other but situated on the same meridian; specifically, we shall take the observations obtained in Vardø (Norway) and Papeete (Tahiti) as they are two of the most distant places for which data are available. By using some of the observation values of those taken during the transit, we shall calculate the Earth-Sun distance.

Both of these expeditions were organised by English scientists. One of them headed for the South Seas to observe the transit in Tahiti, islands that had been discovered two years previously. These observations were carried out by Charles Green and a lieutenant named James Cook, at that time unknown, but who later became very famous on account of his expeditions and adventures. The other team was formed by the director of the Vienna Observatory, Father Maximilian Hell, together with the Danish astronomer C. Horrebow and the young botanist D. Borgrewing, who headed for Vardø in the far north-east of Norway, where they were able to observe

the transit thanks to the famous midnight sun phenomenon. This situation allowed data to be obtained from observers located on the same meridian but in positions very distant from each other.

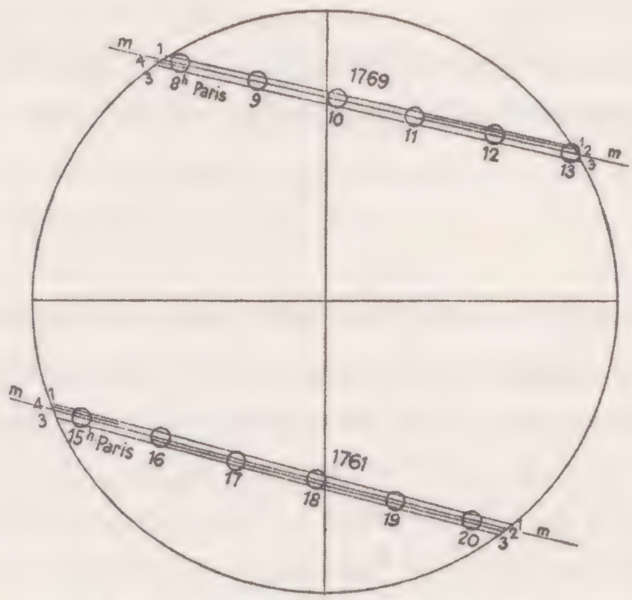
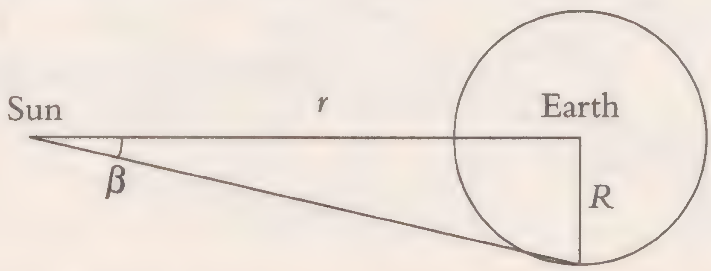


Fig. 29. Transit of Venus across the solar disc
m = track of the planet by its central point
1761: 1. Rodrigues 2. Paris 3. Tobolsk 4. Tahiti
1769: 1. Tahiti 2. Batavia 3. Vardö 4. Paris

Data from the 1761 and 1769 observations of the transit of Venus, published in A History of Astronomy *by A. Pannekoek.*

As was explained further above, parallax is used to measure distances by using angles and a reference distance. By observing the movement of Venus in front of the solar disk, the parallaxes of Venus and the Sun can be obtained, and the distance from the Earth to the Sun can be calculated. To do so, the simplest way is to observe the transit from two places on the terrestrial surface that are sufficiently far away from each other, and, by taking the transit times of both, to deduce the said parallaxes, and finally, to deduce the Earth-Sun distance.



β is the solar parallax, the angle under which the radius of the Earth can be seen from the Sun.

The sun's parallax is angle β shown in the previous diagram. By applying the definition of tangent,

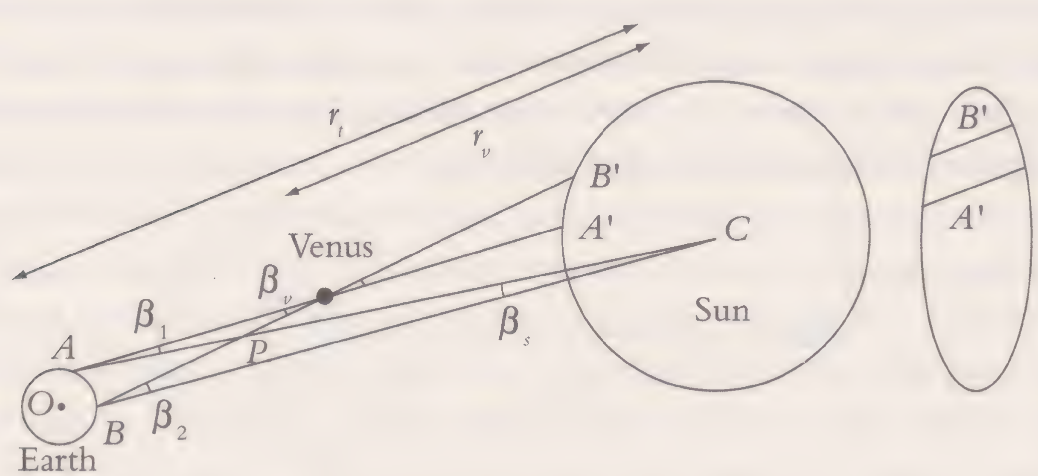
$$\tan\beta = \frac{R}{r},$$

but as the angle is very small, the tangent can be approximated by the angle itself in radians. By simplifying, the Earth-Sun distance, r, verifies that:

$$r = \frac{R}{\beta}.$$

To observe this parallax we would have to be situated on the Sun which, obviously, is impossible and, naturally, for observations taken of the transit of Venus, that does not happen. The observers will be located at different points on the terrestrial surface and will be looking at the Sun from the Earth. In the case proposed above, the location of each of the two observers will be analogous to the situation of each of our two eyes. This phenomenon was exactly the one that was perceived by the two observation teams of the 1769 transit in Tahiti and Vardø.

So we consider two observers situated at locations A and B on the same meridian (so as to simplify the problem's geometry), but at different latitudes. Venus is seen as a point (or small circle) on the solar disk at two different positions, A' and B', due to the fact that the lines of sight from A and B towards Venus are not identical. If we compare the two observations together (see diagram below), it is possible to measure the displacement: the distance $A'B'$ corresponds to the distance between the two positions of Venus observed simultaneously from A and B.



The geometric problem of the Venus transit.

If Venus' movement is observed through the whole transit, a line following its positions throughout that time can be drawn. If we observe from points A and B the result is two parallel lines, one for each place. The distance between both is the displacement of the parallax $\Delta\beta$, which always corresponds to distance $A'B'$. To simplify the mathematical expansion, we can consider the centres of the Earth O, Venus V and the Sun C as coplanar, and we take the observations from two places A and B on the terrestrial surface on that same plane. The triangles APV and BPC have the same external angles at P; consequently, as the sum of the angles of a triangle has to be 180°, it follows that:

$$\beta_V + \beta_1 = \beta_S + \beta_2.$$

By introducing the angle $\Delta\beta$, which measures the distance between two different positions on the trajectory of Venus over the solar diameter ('at each moment', it equals $A'B'$), and by rearranging, we have:

$$\Delta\beta = \beta_2 - \beta_1 = \beta_V - \beta_S = \beta_S\left(\frac{\beta_V}{\beta_S} - 1\right).$$

From the definition, the parallax of Venus is:

$$\beta_V = \frac{AB}{r_t - r_v},$$

and the solar parallax is

$$\beta_S = \frac{AB}{r_t},$$

substituted into the previous equation they result in:

$$\Delta\beta = \beta_S\left(\frac{r_v}{r_t - r_v}\right).$$

Specifically, the solar parallax β_S is:

$$\beta_S = \Delta\beta\left(\frac{r_t}{r_v} - 1\right),$$

where $\Delta\beta$ is the distance between the two lines of the trajectory of Venus seen from

the different observation points, and the quotient r_t/r_v can be obtained by using Kepler's third law. So the cube of that quotient must be proportional to the square of the quotient of the periods of revolution. As the revolution periods of Venus and the Earth are known, 224.7 and 365.25 days, respectively, the solar parallax β_S verifies:

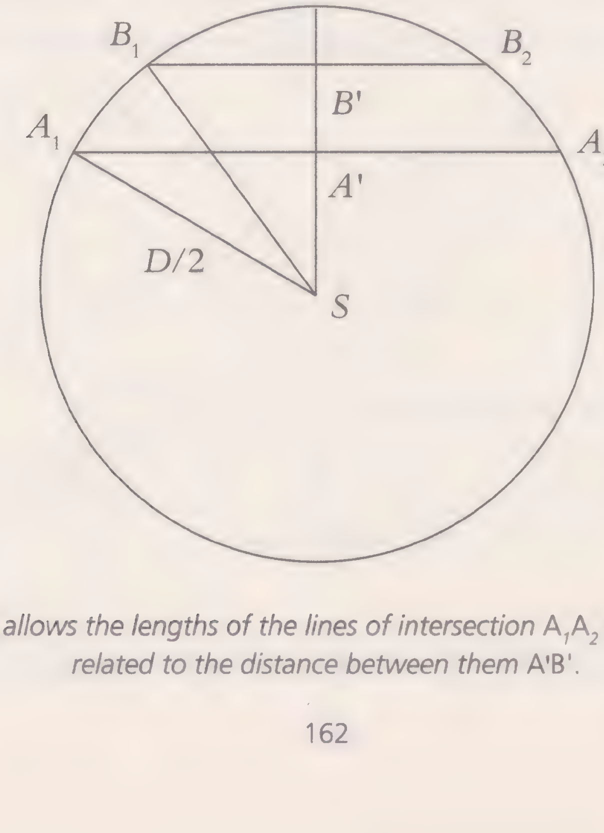
$$\beta_S = 0.38248\ \Delta\beta.$$

$\Delta\beta$ is obtained from the observation data from two places A and B on the same meridian. In this case we shall use the drawing made in the 18th century of the trajectory of Venus from the different locations, all on the same meridian, when the planet's shadow crossed the solar disk. $\Delta\beta$ can be determined in different ways:

1. The simplest is by direct measurement on the upper image on page 159: it is enough to consider a proportion between the diameter D of the Sun on the image and the Sun's angular diameter, which is well known, of 30 minutes of arc expressed in radians:

$$\Delta\beta = \frac{\pi A'B'}{360D}.$$

2. It can also be determined by measuring the lines of intersection in the image. This way increases accuracy, as it is always easier to measure the length of the lines A_1A_2 and B_1B_2 than the distance between the lines $A'B'$.



The figure allows the lengths of the lines of intersection A_1A_2 and B_1B_2 to be related to the distance between them $A'B'$.

By using the Pythagoras triangle theorem $SB'B_1$ and SAA_1 it is deduced that

$$\Delta\beta = A'B' = B'S - A'S = \frac{1}{2}((D^2 - (B_1B_2)^2)^{\frac{1}{2}} - (D^2 - (A_1A_2)^2)^{\frac{1}{2}}).$$

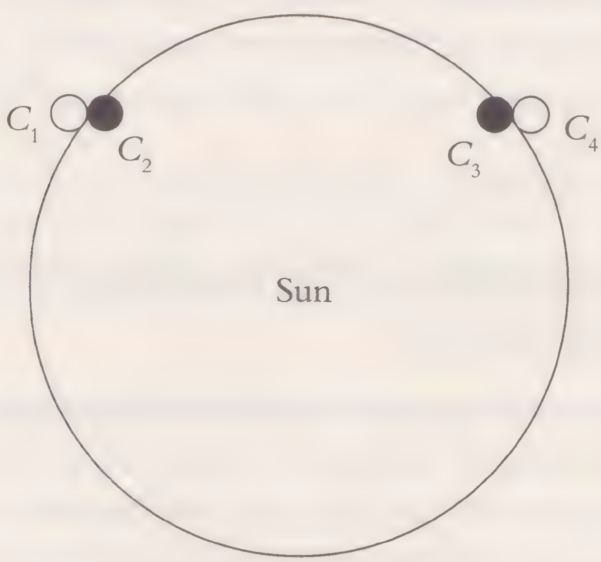
3. Times can be used instead of distances. It suffices to consider the proportion between

$$\frac{A_1A_2}{t_A} = \frac{B_1B_2}{t_B} = \frac{D}{t_o},$$

where t_A and t_B are the lengths of times of the transits A_1A_2 and B_1B_2, and by introducing t_0 as the hypothetical length of transit over the diameter of the solar disk, and with t' as the time corresponding to $\Delta\beta$, it is possible to establish the relation:

$$\frac{t'}{t_0} = \frac{1}{2}\left(\left(1 - \left(\frac{t_B}{t_0}\right)^2\right)^{\frac{1}{2}} - \left(1 - \left(\frac{t_A}{t_0}\right)^2\right)^{\frac{1}{2}}\right).$$

Care has to be taken when using the observation times. As the following figure shows, it should be pointed out that there are external times (C_1 and C_4) and internal times (C_2 and C_3) of observation of the contact ofVenus with the solar edge. Although the black drop effect produces distortions, the internal contacts are always better determined. For this reason, when using numeric values only these are used.



The most precise times are those for the interior contact C_2 and C_3. For this reason, they are the ones used in calculations.

By using the observation data of the 1769 transit for Vardø and Papeete, the following values are deduced, using the methods mentioned and taking distance AB as 11,425 km in a straight line:

Method used	$\Delta\beta$	*Earth-Sun distance* (in 10^6 km)
Direct	0.00019	157
Intersecting lines	0.00020	149
Times	0.00027	144

Results obtained for the Earth-Sun distance, ie, the astronomical unit of distance, according to the three methods described.

As can be seen, the results obtained are reasonable considering the methods used. Nowadays the Earth-Sun distance is taken as $149.6 \cdot 10^6$ km, the definition of the astronomical unit of distance. It should be noted that the second result obtained by using the lines method produces fewer measurement errors than the first method because more precision is obtained by measuring the lines than by measuring $\Delta\beta$ directly. The final method, the one in which the times have been used, is interesting because it provides a clearer analogy with the methods used nowadays for reading times, but it produces more errors through having to make a supplementary hypothesis by supposing that the speed of Venus' movement is constant over the solar surface throughout the transit.

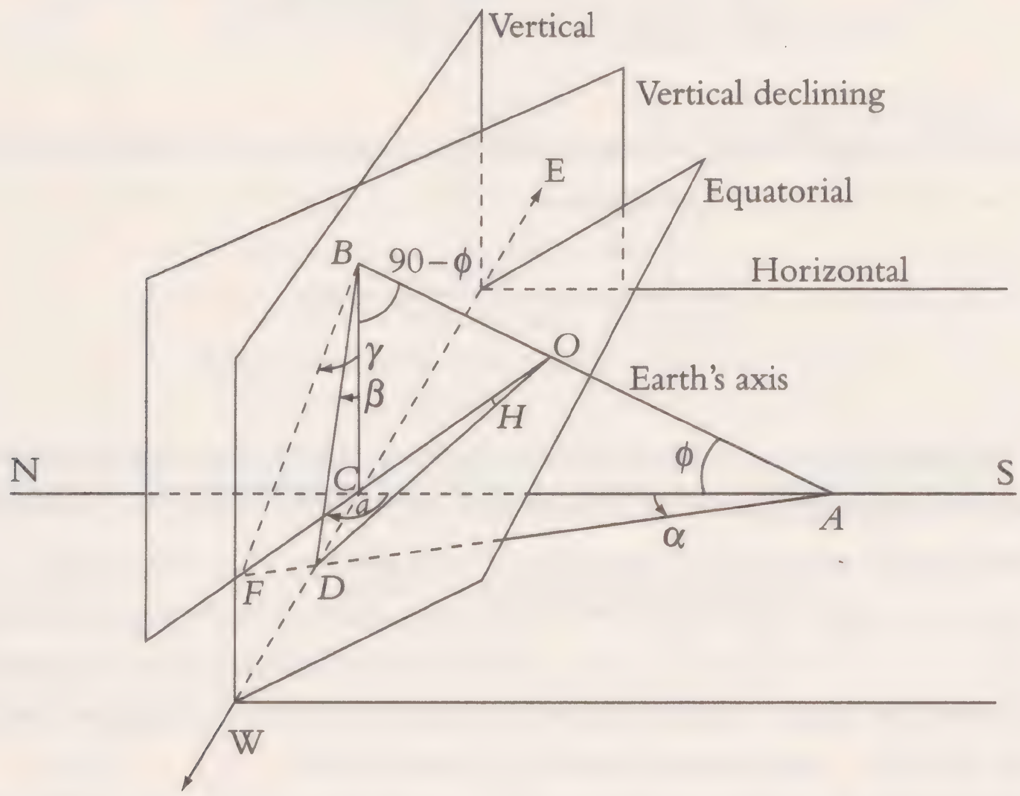
The value of the Earth-Sun distance determined in the 18th century was between 145 and 155 million kilometres (according to the observers). The fact is that that result was improved on at the 19th century transits, but it was calculated more exactly by the use of radar in the year 2000. At present it is considered to be $149.597870691 \cdot 10^6$ km.

Chapter 4: Determination of the hour-lines for a vertical declining sundial

Vertical sundials are usually set on the wall of a building. They are vertical sundials, but unless the wall is oriented in the east-west direction, they are usually oriented towards the horizon over which the Sun moves throughout the day. That will not necessarily be in an east-west direction, but will more likely form an angle with it. To be able to draw hour-lines of a non-oriented vertical sundial, ie those called declining, the angle at which the wall is oriented needs to be known. Later on it

will be explained how that angle a – the wall's azimuth – can be calculated. Let's suppose, to start with, that the angle a is known.

The hour-lines are deduced as in the sundials covered previously, that is, by projecting the lines of an equatorial, horizontal or vertical oriented sundial onto the plane of the vertical declining by using diverse triangles, as shown in the figure. It is useful to remember that 12:00 hours of any vertical sundial coincide with the direction of a plumb line if we place the plumb line at the tip of the gnomon's contact. For the vertical declining plane, the gnomon is, as always, in the direction of the terrestrial rotation axis, ie, Earth's axis.



By projecting the hour-lines of the equatorial clock onto the vertical declining plane, it is deduced that the cotan $\gamma = \sin a \cot \mathrm{H} \sec \phi - \cos a \tan \phi$, where H can be 15 ° and then γ gives us the angle for the 11:00 and 13:00 line, or alternatively H can be 30° and then γ gives us the 10:00 and the 14:00 line, and so on up to 6:00 and 18:00.

By applying the sine formula to triangle CFA it is deduced that:

$$\frac{CF}{\sin\alpha} = \frac{AC}{\sin(180-(a-\alpha))},$$

with $\sin(180-(a-\alpha)) = \sin(a-\alpha)$, taking into account that a runs counter-wise to α, and by using $\sin(a-\alpha) = \sin a \cos\alpha - \cos a \sin\alpha$, it is deduced that:

$$\frac{AC}{CF} = \frac{\sin a \cos \alpha - \cos a \sin \alpha}{\sin \alpha}.$$

But in the triangle ABC, determined by the Earth's axis, it is deduced that $\tan \phi = BC/AC$ and in triangle BFC of the plane of the vertical declining sundial we get $\tan \gamma = CF/BC$, and by simplifying, $\tan \gamma \tan \phi = CF/AC$, where by substituting we deduce the equation:

$$\frac{1}{\tan \gamma \tan \phi} = \frac{\sin a \cos \alpha - \cos a \sin \alpha}{\sin \alpha}.$$

As was concluded above for the horizontal sundial, $\tan \alpha = \tan H \sin \phi$ we can get:

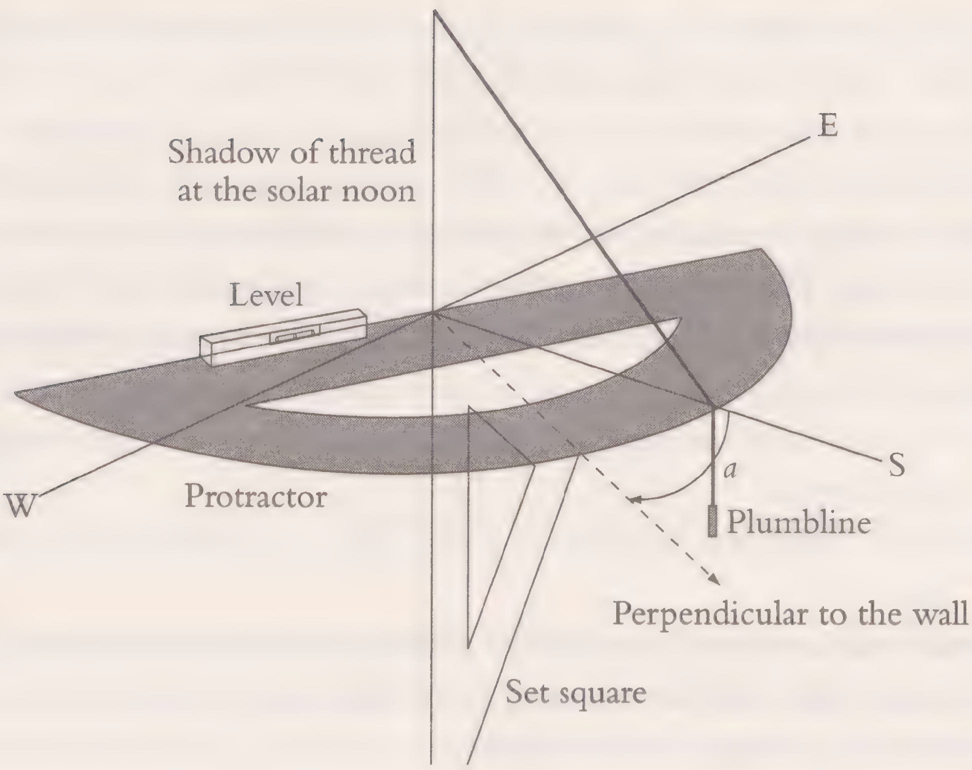
$$\frac{1}{\tan \gamma \tan \phi} = \frac{\sin a}{\tan H \ \sin \phi} - \cos a,$$

where multiplying by $\tan \phi$ the formula is deduced which gives the hour-lines of the vertical declining sundial:

$$\operatorname{cotan} \gamma = \sin a \operatorname{cotan} H \sec \phi - \cos a \ \tan \phi,$$

with γ being the angle between the 12:00 hour-line and another hour-line, and $H = 15°, 30°, 45°\ldots$, respectively, as in the diagram above.

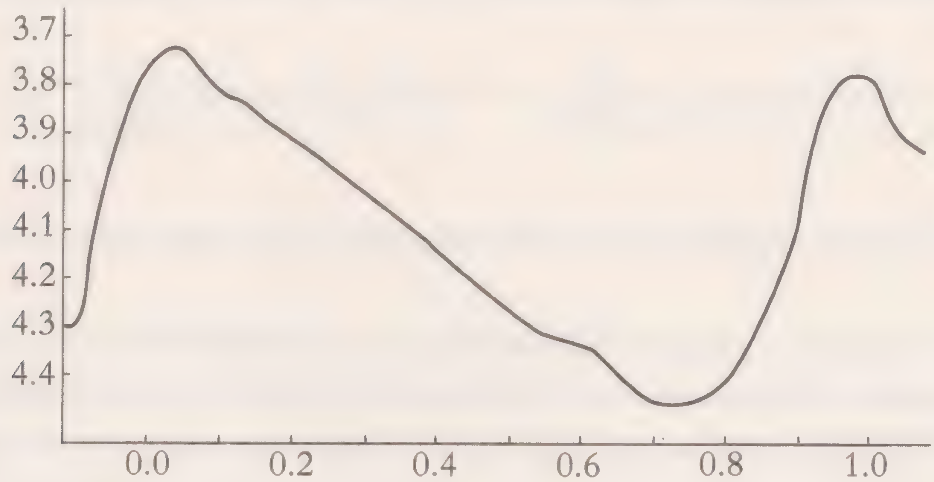
To determine the wall's azimuth, a nail can be hammered into the wall, a string with a weight on the other end tied to the nail, and a spirit level, a set square and a protractor placed in such a way that the protractor is placed on the wall as in the diagram on the next page. The measurement has to be taken at the very instant of the solar midday at the wall's location. The azimuth a of the wall is the angle from the direction south to the direction of the perpendicular to the wall. It has to be remembered that when the Sun passes over the location's meridian (north-south direction), the shadow of the string must be perfectly vertical.



Determination of the azimuth angle a of a vertical wall.

Chapter 5: Determining the light curve of a variable star

For the light curve of a variable star to be drawn, a great many observations are needed. Each point on the light curve has two components (p, m), where p is the phase and m is the apparent magnitude.



The light curve of δ-Cephei.

For each observation, the apparent magnitude has to be assigned by comparing the variable star with two reference stars, *A* and *B*; this is what is known as Argelander's method, which consists of interpolating the observed magnitude values with the reference ones. Obviously, the reference stars must not be variable and their magnitudes must be accurately known, and it is useful if they are the same colour as the variable star. Their apparent magnitudes are m_A and m_B, with $m_A > m_B$, and we denote *Aa v bB*, where *a, b* = 1, 2, 3, 4 or 5, in accordance with the following rules:

- *A*1: We have some doubts about the brightness of *A* and of the variable star (quasi-equals).
- *A*2: Some doubts but in the end we note that *A* is brighter than the variable star.
- *A*3: Both are comparable, but we can clearly see that *A* is brighter.
- *A*4: Right from the start, we note that *A* is brighter.
- *A*5: Star *A* is, without doubt, brighter.
- 1*B*: We have some doubts about the brightness of *B* and the variable star (quasi-equals).
- 2*B*: Some doubts, but in the end we note that *B* is less bright than the variable.
- 3*B*: Both are comparable, but we can clearly see that *B* is less bright.
- 4*B*: Right from the start we note that *B* is less bright.
- 5*B*: Star *B* is, without doubt, less bright.

By using these rules, *a* and *b* can be determined for each observation, and thus the variable star's apparent magnitude can be calculated with the formula:

$$\iota = m_A - (m_A - m_B)\frac{a}{(a+b)} \qquad \text{or} \qquad m = m_B - (m_A - m_B)\frac{b}{(a+b)},$$

and so we get the magnitude, the first component of the point (*m, p*) of the light curve.

To get the second component it is necessary to calculate phase *p* of the variable star at the time of the observation. It is determined from the day, hour and minute of the observation, expressed in Julian days *D*. The ephemerides *E* which gives us the moment of the maxima of a studied star, is also given in Julian days. We also need to know the period *P* which corresponds to the variations in the star's brightness.

If we calculate

$$\frac{(D-E)}{P},$$

we have a decimal number. The 'integer part' of the formula tells us the number of maxima that have taken place since the ephemerides E to the observation D, a result that is not useful for drawing the light curve. What is needed is the 'decimal part', that is, the phase of the variable star at the exact moment of observation:

$$p = \text{decimal part of } \left(\frac{D-E}{P}\right).$$

The Julian day is a day corresponding to any date, but counted in a continuous form. Due to the many reforms made to the calendar and due to other circumstances, such as the non-existence of a year zero, or the existence of leap-years, it was not easy to count the number of days that had gone by between two events. Pope Gregory XIII, for instance, eliminated 10 days: 4 October 1582, was followed by 15 October of the same year. It is plain to see, then, that measuring long periods of time could be complex. So, in 1582 Joseph Scaliger established a system of continuous counting of days, beginning on 1 January of the year 4713 BC at 12 noon (in those days the day began at midday, when the Sun passed over the location's meridian, and not as it is now, when the day begins at midnight) and he counted the days correlatively; this number is called the Julian number. For example, 1 January 2014, at midday, is Julian day 2,456,659.

Bibliography

Cox, B., Cohen, A., *Wonders of the Universe*, London, Harper Collins, 2011

Moore, Sir P., *Guide to the Night Sky: A Guided Tour of the Stars and Constellations*, London, Philips/Octopus Publishing Group, 2005.

Ros, R.M., "Astronomy and Mathematics", *Teaching and Learning Astronomy*, Cambridge University Press, 2005.

– : "The Transit of Venus: an Opportunity to Promote Astronomy", *Teaching and Communicating Astronomy*, Granada, EDP Sciences, 2005.

Ridpath, I., Tirion, W., *Collins Stars and Planets Guide*, London, 2011.

–: *The Monthly Sky Guide*, Cambridge University Press, 2009

Sagan, C., *Cosmos*, London, Abacus, 1995.

Savoie, D., *Sundials: Design, Construction and Use*, New York, Praxis, 2009

Index

Cosmic Calculations

Astronomy and mathematics

Mathematics and astronomy began their journey together thousands of years ago. The Babylonian, Greek and Hindu civilisations were among the most passionate in their quest to establish the rules – halfway between magic and science – for the cosmic dance of the stars and planets. Because it was such a useful guide for travellers, astronomy was one of the first and most important applications that humankind found for scientific thinking. From those early days, mathematics has been the vehicle that has helped the intellect to travel even further – to the edges of the universe and beyond.